

doi:10.3799/dqkx.2025.086

结合短时傅里叶变换与注意力的震相识别模型

雷鸣^{1,2}, 周云耀^{1,2,3,4*}, 向涯^{1,2,3}, 吕永清^{1,2,3}

1 中国地震局武汉地球观测研究所, 湖北武汉 430071

2 中国地震局地震研究所, 湖北武汉 430071

3 中国地震局地震大地测量重点实验室, 湖北武汉 430071

4 武汉引力与固体潮国家野外科学观测研究站, 湖北武汉 430071

摘要: 地震震相拾取的准确性直接影响震源定位和震级估计的精度, 然而传统方法对复杂地震信号的特征捕捉能力有限。本文提出了一种融合多尺度注意力机制和短时傅里叶变换的双分支模型, 该模型通过两个分支分别捕获信号的时间特征和时频特征, 并结合注意力机制实现多尺度的特征增强。实验结果表明, 在 100ms 的误差范围内 P 波震相拾取的识别精度和召回率分别达到了 95.69% 和 88.97%, S 波震相拾取的识别精度和召回率分别达到了 87.98% 和 77.25%。P 波的到时误差均值和标准差分别达到了 18.76ms 和 27.13ms, S 波的到时误差均值和标准差分别达到了 25.97ms 和 36.14ms。同时模型的参数量(Parameters)仅有 0.35 M, 计算开销(FLOPs)为 71.38 M。与同类模型相比, SEN 模型不仅在性能上取得显著提升, 同时在参数量和计算开销上具有一定优势, 为地震监测的实时应用提供了有力的技术支持。

关键词: 震相识别; 深度学习; 卷积神经网络; 注意力机制; 时序

中图分类号: P315

收稿日期: 2025-02-17

A Phase Picking Model Integrating Short-Time Fourier Transform and Multi-Scale Attention

LEI Ming¹, ZHOU Yunyao^{1,2,3,4*}, XIANG ya^{1,2,3}, LV Yongqing^{1,2,3}

1 Wuhan Institute of Earth Observation, CEA, Wuhan 430071, China

2 Institute of Seismology, CEA, Wuhan 430071, China

3 Key Laboratory of Earthquake Geodesy, CEA, Wuhan 430071, China

4 Wuhan Gravitation and Solid Earth Tides, National Observation and Research Station, Wuhan 430071, China

基金项目: 国家重点研发计划(2022YFC3003804)

作者简介: 雷鸣(1999-), 男, 研究生, 主要从事地震预警系统研究。E-mail: leiming22@mails.ucas.ac.cn, ORCID: 0009-0006-8994-9942

*通讯作者: 周云耀, 男, 研究员, 主要从事地震观测技术与仪器研究, E-mail: joewhcn@126.com.

Abstract: Seismic phase picking is a critical task in earthquake monitoring, as its accuracy directly impacts the precision of hypocenter localization and magnitude estimation. However, traditional methods are often limited in their ability to capture the characteristics of complex seismic signals. This study proposes a dual-branch deep learning model that integrates a multi-scale attention mechanism and short-time Fourier transform (STFT). The model extracts temporal features through a time-domain branch and captures time-frequency representations via a frequency-domain branch, while leveraging the attention mechanism to enhance multi-scale features. Experimental results show that within a 100 ms error threshold, the proposed model achieves a P-wave picking precision and recall of 95.69% and 88.97%, and an S-wave precision and recall of 87.98% and 77.25%, respectively. The mean and standard deviation of arrival time error for the P-wave are 18.76 ms and 27.13 ms, while for the S-wave they are 25.97 ms and 36.14 ms. Moreover, the model contains only 0.35M parameters and incurs a computational cost of 71.38M FLOPs. Compared with existing models, the SEN model not only achieves competitive performance but also demonstrates advantages in model size and computational efficiency, offering strong potential for real-time seismic monitoring applications.

Key words: seismic phase picking; deep learning; convolutional neural network; attention mechanism; time series

深度学习作为一种高效且广泛应用的方法,已经在多个领域取得了显著的成果(LeCun et al., 2015)。在地震监测领域,随着数据量的迅速增长,处理工具更加高效、健壮的需求日益增加,使得深度学习成为地震学领域学习和应用的热点。目前,地震信号的检测和震相拾取是深度学习在地震学领域中的主要研究方向(Bergen et al., 2019; Mousavi and Beroza, 2022)。

强噪声下地震信号检测和震相拾取是地震监测中具有挑战性的问题。地震信号检测是指在地震传感器记录的各种非地震信号和噪声中识别出地震事件;震相拾取则是测量地震信号中不同震相(如 P 波和 S 波)的到时。尽管这两项任务极为相似,但目标并不完全相同。对于检测,主要目标是最小化漏检率和误检率;而在震相拾取中,除了

保证检测的准确性外,还需要尽可能地提高到时的拾取精度。这将为后续的地震定位等任务提供有力支持,对于地震定位而言,其对地震到时测量极为敏感,P 波到时 10 毫秒的误差可能导致定位上出现数十米的偏差(Mousavi et al., 2020)。

深度学习以一种新的方式学习地震相位的波形特征,不再显式地定义特征函数,而是通过监督学习从大量标记了相位的波形中自动学习特征。这类方法完全由数据驱动,减少了对主观参数设置的依赖。此外,深度学习模型可以提供可靠的震相类型判别(何彬等., 2024; Zhou et al., 2019),这对地震定位非常重要,但在传统方法中通常难以准确获得。

目前,有不少学者在震相识别方法上

已经提出了非常优秀的模型。Ross(2018)提出了基于卷积神经网络(CNN)的震相检测器,该模型通过一系列卷积层提取地震波形的时空特征,能够同时检测 P 波和 S 波,其网络采用多层卷积设计,使其能够提取不同尺度的特征,且在南加州地震数据中的实验结果表现优异。Zhu and Beroza(2019)提出了 PhaseNet,这是一种基于 Unet 结构,使用三分量地震波形作为输入,并输出输入序列上每个点的 P 波、S 波和噪声的概率分布。Li et al. (2023)提出了一种基于 UNet++ 改进结构的震相检测模型 CSESnet,对四川-云南地区的地震台网数据进行训练,融合了多尺度特征以增强模型对 P 波的检测能力。四川泸县 M6.0 等实际强震中的测试表明,该模型在强震环境下亦具有良好的泛化能力,表现出较强的实际部署潜力。Wang et al.(2019)开发了一种基于深度学习的地震到达时间拾取模型,采用了 CNN 结构,证明了该模型在不同噪声水平下均具有较高的到时拾取精度。虽然前人的这些研究在方法以及结果上存在一定的差异,但普遍发现基于深度学习的方法比传统算法更具优势(Zhu et al., 2019;Saad and Chen, 2021;Chai et al., 2020;Chen et al., 2019;Zhu et al., 2023;陈国艺等., 2023;胡进军., 2023)。通过对比多种模型结构可以发现,多尺度信息融合可以给地震信号识别在效率和准确性等方面带来显著的提升。

然而,这些方案在捕捉地震信号的长

程依赖性方面仍存在局限。为了解决这一问题,前人引入了注意力机制(Li et al., 2025;Li et al., 2021;Liao et al., 2021;Zhang et al., 2023)。

近年来,深度学习领域的发展都偏向于使用注意力机制。注意力机制最初在 NLP 领域取得成功,能够让模型更关注重要的特征信息(Vaswani et al., 2017)。对于地震信号检测任务,地震信号的数量往往比噪声信号少,因此引入注意力机制可以增强模型对关键信号特征的关注度,提高噪声和地震信号震相的识别能力。例如,Mousavi et al. (2020)提出的 Earthquake Transformer 模型采用了基于自注意力机制的深度学习框架。该模型引入了 Transformer 结构,通过捕捉地震波形的长程依赖性,能够更好地分辨震相与噪声之间的微小差异。实验表明,Earthquake Transformer 能够显著提升地震波的识别精度。Xiao et al.(2021)开发了 Siamese Earthquake Transformer,这是一种双输入的深度学习模型,通过输入相同地震台阵中的两个不同台站的波形数据,结合 Transformer 架构实现了地震检测和相位拾取,并在地震台阵数据上表现出色。Lan et al.(2025)提出的 Cross-Attention 模型是一种基于交叉注意力机制的地震相位拾取模型。该模型利用多头交叉注意力机制捕捉地震波形的关键特征,显著提高了复杂地震环境中的相位拾取准确性。实验表明,该方法在噪声抑制和弱信号检测方面表现优越。

前人的这些相关工作展示了注意力机制在地震检测和震相拾取中的多样化应用，然而传统注意力机制的计算开销过高，为了解决该问题，有学者提出了更加高效的注意力机制。例如，Ouyang et al.(2023)提出了高效多尺度注意力模块(Efficient Multi-Scale Attention, EMA)，通过跨空间学习和并行子网络结构，结合局部和全局的通道交互，捕获短程和长程的依赖关系。该模块结合跨尺度交互机制，极大的降低了计算开销，提升了模型对复杂信号的处理能力。

地震信号是一种时频信息丰富的数据，频域信息对于准确识别震相具有重要意义，而现有研究当中大多数并未充分利用到频域信息。已有研究表明，深度学习在时间序列应用领域中融合频域信息可以显著提升效果。Ye(2024)提出的 ATFNet 模型结合了时频域信息，通过自适应时频融合模块，提升了对复杂时序信号的建模能力。FEDformer 在处理长时间序列数据方面表现突出，其通过频率增强分解模块对时间序列的频率特征进行提取和建模(Zhou et al., 2022)。

这些研究表明，对于时间序列任务而言，频域信息是一个非常重要的部分。前人在时序任务中用到的频域信息往往是基于傅里叶变换的，而傅里叶变换获取到的是一种全局频率分布情况，对于震相识别任务而言并无较大帮助。基于傅里叶变换的短时傅里叶变换 (STFT) 可以解决该问题(Allen,

1977)，STFT 能够提供信号在不同时序切片上的频率分布情况。通过对地震数据的研究发现，STFT 构建的时频图可以较好地描述 P、S 波到时的频域信息。同时，STFT 在时域上的小窗口也可以很好地将频域信息与时域信息结合，丰富模型的输入特征，有助提高对复杂信号的处理能力。

综上所述，现有研究在地震震相识别方面取得了显著进展，但仍存在一些有待解决的问题。大多数研究主要依赖时域特征，对频域信息的利用相对不足，而频域信息往往在识别复杂地震信号和提升抗噪性能方面具有重要作用。其次，现有网络结构通常包含大量可训练参数且计算成本较高，这限制了它们在实时监测和边缘计算设备上的应用。轻量化网络不仅能够降低硬件需求和运行成本，还能够直接部署于低成本的边缘计算设备，实现实时数据监测。为此，本研究提出了一种结合短时傅里叶变换和多尺度高效注意力的轻型震相识别网络 (Short-Time Fourier Transform with Efficient Multi-Scale Attention Phase Picking Network, SEN)，以期望通过结合时域和频域信息从而提升震相识别的性能。该网络通过引入 EMA 注意力模块有效聚焦关键时频特征，并结合 STFT 提取信号的频域信息，以期望增强对复杂地震模式的捕捉能力。与此同时，网络的设计注重轻量化，采用深度可分离卷积等优化策略。

1 模型设计

1.1 多尺度注意力模块

本研究采用 EMA 模块(Ouyang et al., 2023)作为注意力的计算, 该模块通过在通道和空间维度上的交互, 增强模型对特征的敏感性。该模块具有多尺度并行结构, 能够在保持多尺度特征表达的前提下, 有效减少模型深度、降低顺序处理的计算开销。其使用 1×1 卷积和 3×3 卷积构建了两个并行分支, 分别用于获取全局和局部注意力, 模型具体结构如图 1 所示。EMA 模块首先将输入特征按通道进行分组, 以减少计算量。在 1×1 卷积分支中, 首先对分组后的特征图在两个空间维度分别进行全局平均池化操作, 生成一个包含空间信息的特征图。然后, 将该特征图通过一个共享的 1×1 卷积层, 用以

提取局部的空间信息, 生成空间注意力矩阵, 该矩阵用于描述特定位置的重要性。随后, 再对 1×1 卷积分支的输出进行一次全局池化再通过归一化指数函数 (softmax) 将向量转化为概率分布, 生成通道注意力权重, 完成对不同通道的重要性调整, 再将该通道注意力作用于 3×3 卷积的结果, 实现跨空间学习。在 3×3 卷积分支中, 首先通过 3×3 卷积捕获输入特征的空间局部信息, 得到空间特征图。然后, 对该特征图进行全局池化, 将空间信息压缩到通道维度上, 生成通道注意力矩阵再将其作用于 1×1 卷积分支的结果。最终, 将两个分支生成的注意力矩阵相加, 并通过 S 形激活函数 (sigmoid) 生成多尺度的全局注意力矩阵, 作用于原始输入, 实现对关键通道和空间位置的多尺度增强。

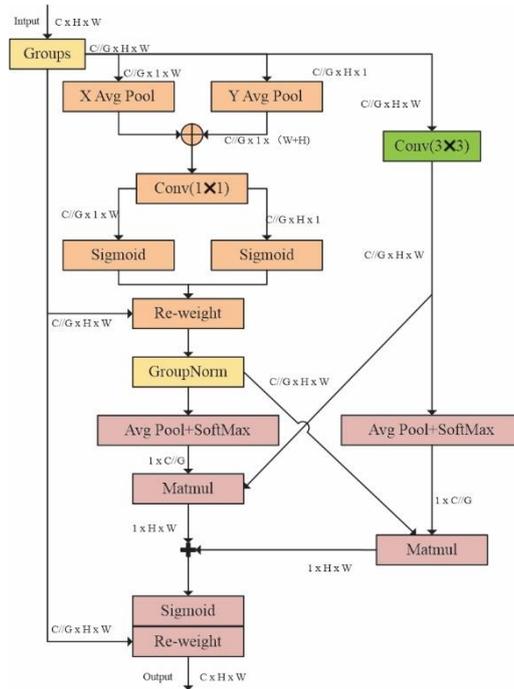


图 1 EMA 模块结构图

Figure 1 Structural of the EMA Module

相比于其他复杂的注意力机制, EMA 模块通过多尺度并行结构和轻量级的卷积

操作，显著降低了计算开销。这对于处理高维度的地震信号数据尤为重要，可以保证模型的实时性和可扩展性。同时提供了跨空间、跨通道的多尺度注意力，这种多尺度的注意力可以显著提高对于地震信号这类复杂信号模式的感知能力。

1.2 短时傅里叶变换模块

目前的引入频域信息的相关工作中往往都是基于离散傅里叶(DFT)或快速傅里叶变换(FFT)以及其变种以获取到频率信息，如前文所提到的 ATFNet 等。但这种方式获取到的是输入的全局频域信息，而对于地震信号而言，我们往往更需要关注到某一段时间的频率分布情况，而短时傅里叶变换(STFT)便是将时间切分为若干个时间切片，随后对每个切片计算得到它们的频谱图，计算方法如式 1 所示

$$STFT(t, f) = \int_{-\infty}^{+\infty} [h(u)g(u-t)] e^{-j2\pi fu} du \quad (1)$$

式中 $g(u-t)$ 代表移位窗函数， $h(u)$ 为原始信号。但 STFT 受限于时间窗口的大小限制，当窗口较小时，虽然时间分辨率提高了，但频率分辨率较低，为了兼顾时间分辨率以及频率分辨率本次研究中窗口大小设置为 64，再通过最邻近插值的方式对时频图进行上采样以匹配张量的尺寸。

1.3 整体模型设计

本文模型的整体模型结构如图 2 所示。该模型采用双分支架构，将地震信号的时域和频域信息融为一体，以提升震相识别的准确度和鲁棒性。具体而言，模型分为处理原始波形的时域分支与利用 STFT 提取频域特征的频域分支。两个分支在各自的空间中对信号特征进行提取，并通过注意力机制突出关键特征，随后在通道维度上融合。最终，融合特征通过两个卷积层输出震相分类结果与时间偏移的回归预测。

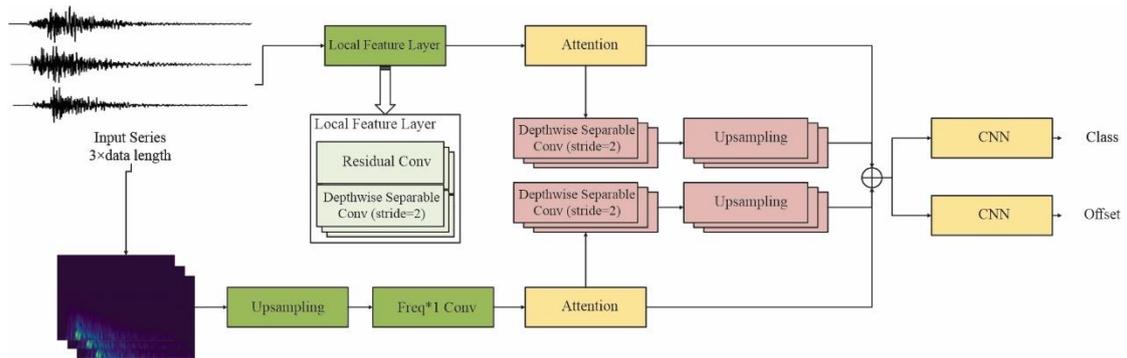


图 2 SEN 模型结构图

Figure 2 Structural of the SEN Module

为有效降低计算成本并保持特征提取的有效性，模型中所有卷积运算均采用深度可分离卷积。该卷积方式将标准卷积分解为

深度卷积与逐点卷积两个步骤，由此显著减少参数量与计算开销。在频域分支中，首先对三通道的地震波形序列进行 STFT，获取

相应的时频图，STFT 能够捕获信号在时间和频率上的局部变化特征，为识别瞬态震相特征提供有力支持；之后，对得到的时频图进行上采样以统一特征尺度，并在频率维度上执行 $n \times 1$ 卷积，以提取每个时间步对应的频率特征；紧接着，通过注意力模块对频率特征进行加权，使模型更关注对震相识别具有重要贡献的位置；最后，频域分支通过编码-解码结构扩大感受野并提取时频空间的全局特征，为后续与时域信息的融合做好准备。在时域分支中，首先利用残差卷积加步长为2的深度可分离卷积对原始三通道时序信号进行特征提取，以便从信号的时间序列中捕获局部特征；随后，通过注意力模块对特征进行加权，使模型在时间维度上聚焦于更重要的位置，从而提升对关键震相位置的辨识能力；最后，该分支同样利用编码-解码结构扩大感受野并捕捉全局时序特征。融合阶段，将频域与时域两个分支的输出及其局部关键特征在通道维度上进行整合。这种多模态融合主要利用时域信息对信号形态和振幅变化的敏感性以及频域信息对频谱结构的洞察力获得更全面、更具判别力的特征表示。经过融合后，模型最终通过两个标准卷积层生成震相类别和补偿值预测，从而实现了对震相类型与到时偏差的高精度估计。

总体而言，该双分支架构整合了时域和频域特征，辅以注意力机制和多尺度特征提取策略，在保持较低计算成本的前提下，显著增强了震相识别的精度与稳定性，为后续地震事件分析和定位奠定了坚实的基础。

2 实验结果与分析

2.1 数据集设计

本文研究的数据集来源于 Mousavi(2019)公开的 STanford EArthquake Dataset (STEAD)，从中随机抽取了 10 万条纯噪声记录以及 20 万条地震记录。地震数据为采样率为 100Hz 的三分量（EW、NS、UD）数据，其中还包含 P、S 波到达时间。地震事件的震中距分布范围较广，数据信噪比（SNR）差别较大，在随机抽取的数据中，地震事件数据的信噪比以及震中距情况如图 3 所示。从图 3 中可以看到，数据整体的信噪比较低，在这种高噪声水平下地震事件震相的准确识别具有较大挑战，但这也更符合真实情况，训练出来的模型也具有更好的鲁棒性。

随机抽取数据的观测设备及地震事件的分布如图 4 所示，图中红点为观测设备位置，黑点为地震事件位置。由图 4 可知，观测设备和地震事件的分布范围较广，基本涵盖各类地质条件。

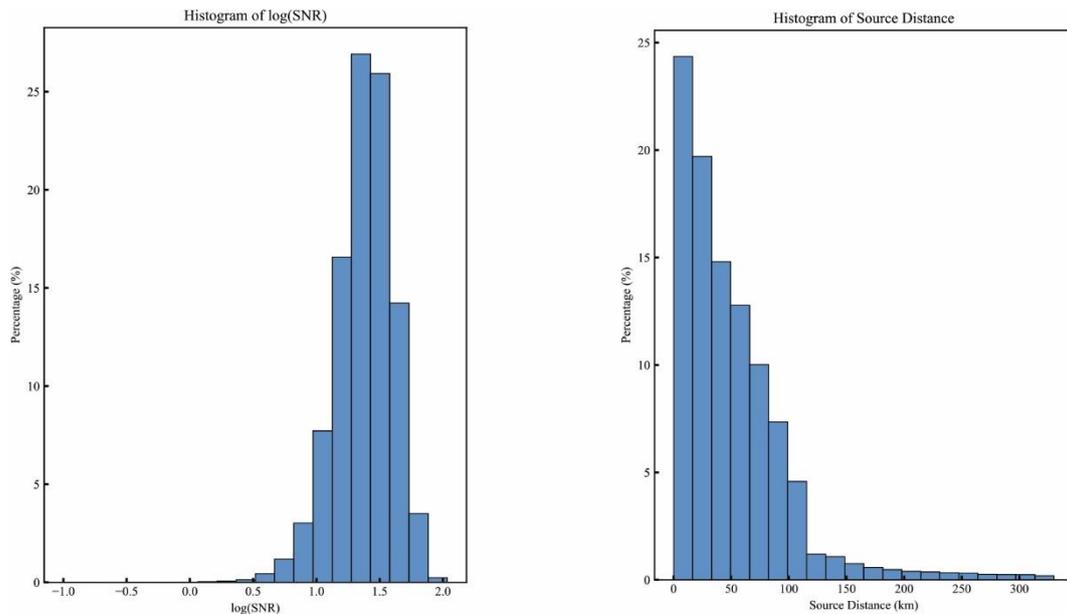


图3 数据集信噪比及震中距分布

Figure 3 Distribution of $\log(\text{SNR})$ and Epicentral Distance of the Dataset

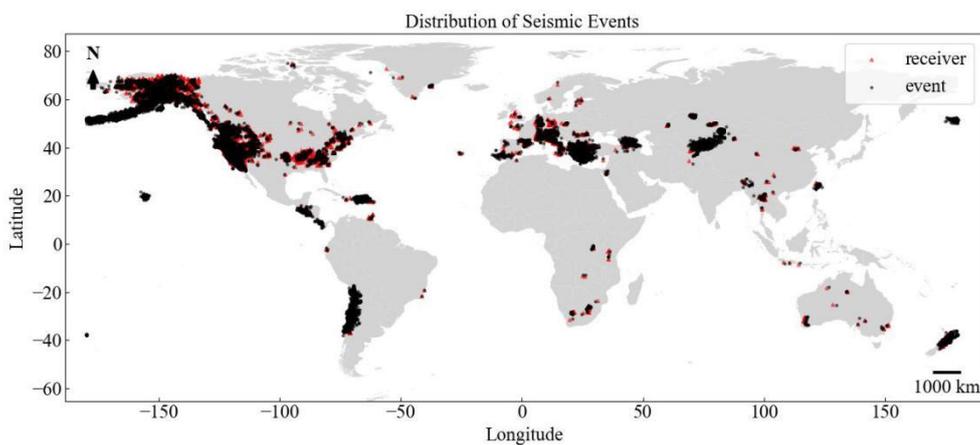


图4 地震事件分布图

Figure 4 Distribution of Seismic Events

在数据预处理方面，为了增强模型的泛化能力以应对更复杂的噪声环境，首先对所有数据样本添加了随机噪声，随后进行了标准化处理，并引入了随机平移操作。

对于每一条非地震记录，截取数据的起点随机，若终点超出了数据长度，则补零，对于地震记录而言，每一条输入数据至少包含一个 P 波到时或一个 S 波到时。首先随机选择 S 波或 P 波作为目标点，每一个目标点

必须包含到当前输入中，在保证目标点包含在当前输入的情况下，加入随机平移，即该输入的数据起点在目标点前 $n(n < 2048)$ 个点中随机选择，若起止点超出的数据记录长度或序列长度不足 2048 点时则对超出部分或不足部分补 0。通过上述预处理步骤，生成的输入数据不仅包含必要的地震信号特征，还具备一定的随机性和鲁棒性，以提升模型的泛化能力。

在本次研究当中，对于标签的设计，将 2048 个点按照步长为 8 的方式缩小，即输出和标签长度为 256，从而直接减少计算开销。但对于每一个目标点而言不一定都可以被 8 整除，因此为了获取到准确的到时位置需要添加一个补偿，用于描述缩小尺寸后的到时点与真实到时点之间的距离。因此模型标签包含分类标签以及补偿标签，分类标签

$$label_{class}(i) = \begin{cases} 1 & i = \left\lfloor \frac{P_{idx} - Start_{idx}}{8} \right\rfloor \\ 2 & i = \left\lfloor \frac{S_{idx} - Start_{idx}}{8} \right\rfloor \\ -1 & i = \left\lfloor \frac{P_{idx} - Start_{idx}}{8} \right\rfloor - 1 \text{ or } \left\lfloor \frac{S_{idx} - Start_{idx}}{8} \right\rfloor - 1 \\ 0 & \text{else} \end{cases} \quad (2)$$

$$label_{offset}(i) = \begin{cases} (P_{idx} - Start_{idx}) \bmod 8 & i = \left\lfloor \frac{P_{idx} - Start_{idx}}{8} \right\rfloor \\ (S_{idx} - Start_{idx}) \bmod 8 & i = \left\lfloor \frac{S_{idx} - Start_{idx}}{8} \right\rfloor \\ 0 & \text{else} \end{cases} \quad (3)$$

式 2 和式 3 中 $label_{class}$ 代表分类标签， $label_{offset}$ 代表补偿标签， i 代表标签数据中的第 i 个点， P_{idx} 代表地震记录中的 P 波到时点， S_{idx} 代表地震记录中的 S 波到时点， $Start_{idx}$ 为地震记录中的随机起点， \bmod 表示取模运算。

2.2 实验设置

本研究中模型基于 Python 中的 pytorch 库实现，输入尺寸设计为 2048， $batch_size$ (批次大小)设置为 64，优化器采用 adamw 优化器，学习率设置为 0.0001，

$$Focal\ Loss(cx, C) = - \sum_i \alpha_{C_j} \left(1 - \frac{e^{cx_i C_j}}{\sum_k e^{cx_i k}} \right)^{\gamma} \log \left(\frac{e^{cx_i C_j}}{\sum_k e^{cx_i k}} \right) \quad (4)$$

公式 4 中 cx 代表模型输出中每个点被

用于区分噪声、P 波、S 波，补偿标签用于标记当前到时距离与真实到时之间的点数。在分类标签中，将当前输入的震相到时点处设置为类别 1，前后一个点处设置为 -1，即在损失计算中忽略该点的损失。在补偿标签当中，将当前输入的震相到时点设置为距离真实到时点的差值，具体如公式 2 和公式 3 所示。

采用余弦退火算法进行学习率的下降从而避免过拟合。

在损失函数的选择上，考虑到地震波到时点相对稀少，到时数据样本的数量与噪声相比存在显著的不平衡，而焦点损失 (Focal Loss) 能够有效减少易分类样本的损失贡献，增强模型对少数类 (P 波和 S 波) 的学习能力，因此采用 Focal Loss 用于分类的损失计算。对于单个序列中的每一个时间点 $i (i = 1, 2, \dots, n)$ ，焦点损失表示如公式 4 所示。

分类为 C_j 的概率分布， C 代表该输出序列中

每个点的真实分类, k 代表存在 k 个类别, α 代表每个类别的权重。对于 α 权重的设置, 本文根据抽取的样本分布情况设置为 1, 4, 4。 γ 是对于难以识别样本的调节因子, 设置越大则难以识别样本偏差较大时产生的损失就越大, 本文设置与 Focal Loss 的论文一致(Lin et al., 2017)。

对于补偿值的训练是一个回归任务, 故而采用均方误差 (MSE)计算损失, MSE 主要用于衡量预测补偿值与真实补偿值之间的差异。为了只针对关键点进行计算, 引入了一个符号函数 $t(i)$, 当第 i 个点的真实类别为 P 波或 S 波时为 1, 否则为 0。方法如式 5 所示:

$$Offset Loss(tx, T) = \sum_i t(i)(tx_i - T_i)^2 \quad (5)$$

tx_i 代表模型输出中每个可能为震相到时点的补偿值, T 代表该输出序列中每个点的真实补偿值, $t(i)$ 是一个符号函数。最终的损失函数如式 6 所示。

$$Loss(cx, tx, C, T) = Focal Loss(cx, C) + Offset Loss(tx, T) \quad (6)$$

本文采用焦点损失函数(Focal Loss)和均方误差(MSE)损失函数进行组合的方式, 使模型能够在分类任务和回归任务之间实现有效的权衡, 提升整体性能。

本次对比实验中基线模型选择为 LPPNL、EQT 以及 PhaseNet。LPPNL 模型是 Yu and Wang.(2022)于 2022 年提出的一种轻量化震相识别模型, 在保持较小计算开销和模型参数量的同时保证了较高的识

别精度。(Yu et al., 2023)。EQT 模型是 Mousavi(2020)提出的一种基于 Transformer 结构的震相识别模型, 该模型通过结合自注意力机制, 在多种地震数据集上表现出色。PhaseNet 模型是震相识别领域的经典模型, 由 Zhu and Beroza.(2019)提出的一种基于 Unet 结构的震相识别模型。为了实现公平对比, 这三个模型的输入尺寸全部调整为了 2048, 其他参数与原模型一致, 对于 LPPNL 的 Stride 参数本文设置为 8。

2.3 实验结果

本次研究中对于到时误差 T_{error} 的计算具体如下, 将模型输出的到时点记为 $T_{predict}$, 实际 P 波或 S 波到时记为 T_{true} , 到时误差 T_{error} 的计算方式如式 7 所示。

$$T_{error} = T_{predict} - T_{true} \quad (7)$$

对于每一个 $|T_{error}| \leq 100ms$ 的样本被认为正确识别, 对于 $|T_{error}| > 100ms$ 的样本则认为被识别为了噪声。

对于震相拾取任务而言, 不仅需要模型能够准确识别出 P 波、S 波等不同震相的类型, 同时要保证到时识别的精确性, 因此评价震相拾取模型需要考虑其分类性能和到时定位的综合能力。对于分类性能的评估, 本文采用常见的分类模型评价标准即精度 (Precision)和召回率(Recall), 精度越接近于 1, 则代表模型识别准确率越高, 召回率越接近于 1, 则代表模型识别的越全面。对于定位能力本文主要考虑 T_{error} 的均值与标准差, 均值越接近于 0, 则代表模型整体的预测偏

差越低；标准差越接近于 0，则代表模型的预测偏差的离散程度较低，即代表模型更加的稳定，这代表在模型输出中出现较大偏差的比例越低。

表 1 模型各项指标
Table 1 Performance of the Model

模型	Precision		Recall	
	P 波	S 波	P 波	S 波
SEN	0.9569	0.8798	0.8897	0.7725
LPPNL	0.9434	0.8815	0.8632	0.7208
EQT	0.9350	0.8734	0.8541	0.7027
PhaesNet	0.8230	0.8782	0.7640	0.7030

四个模型分类性能在验证集上的评估结果如表 1 所示。由表 1 可见，SEN 模型在 P 波的识别上取得了较高的精度和召回率。表明模型对于 P 波事件的识别具有较高的精确度，能够稳定地检出多数 P 波事件。S 波的精度略低于 LPPNL，但召回率指标相对于其他三个对照模型得到了显著提升，表明本文模型对震相特征的提取能力较强，对 S 波的识别相比其他模型更加全面。

表 2 到时误差均值以及标准差

Table 2. Mean and Standard Deviation of Arrival

模型	Time Errors			
	Mean (ms)		Std (ms)	
	P 波	S 波	P 波	S 波
SEN	18.76	26.13	27.13	36.14
LPPNL	18.81	25.97	27.71	35.18
EQT	23.90	30.62	32.76	40.11
PhaseNet	18.51	26.63	28.41	36.81

模型的到时定位能力方面，P 波、S 波到时误差的均值和标准差如表 2 所示。SEN 模型在 P 波和 S 波的到时误差均值方面与 LPPNL、PhaseNet 相近，显著优于 EQT。在

P 波到时误差均值方面，SEN 略高于 PhaseNet，但在 P 波到时误差的标准差上低于 PhaseNet 和 EQT，说明与之相比 SEN 在 P 波到时的稳定性上有所提高，但比 LPPNL 表现略差可能是由于模型在设计上更加关注召回率和精度。

在 S 波到时定位上，SEN 模型的误差均值略高于 LPPNL，略低于 PhaseNet，显著优于 EQT。而在到时误差的标准差上相比与 LPPNL 略高，这可能是由于 SEN 识别到了更多的 S 波导致。但综合分类性能来看，SEN 在 S 波拾取上相比三个对照模型具有显著提升，且在定位能力上能够达到对照模型的水平。

表 3 模型参数量与计算量对比

Table 3. Parameters and FLOPs

模型	Parameters/M	FLOPs/M
SEN	0.35 M	71.38 M
LPPNL	0.66 M	111.98 M
EQT	2.59 M	87.91 M
PhaseNet	0.17 M	17.06 M

四个模型的参数量(Parameters)以及对于单条输入的计算量(FLOPs)如表 3 所示。SEN 模型相比于 LPPNL 和 EQT 的计算量都有所减少，显著大于 PhaseNet，但结合前面的定位与分类性能来看，SEN 的定位与分类性能相比与 PhaseNet 识别性能与误差控制得到了显著提升，显示出较强的综合竞争力。同时从参数量的绝对数量上来讲也能够支持在计算资源受限的应用场景，这对于利用边缘设备对地震进行更精确的实时监控、快速处理大量地震数据的在线系统将具有

重要价值。

表 4 模型运行效率对比

Table 4. Model Runtime Efficiency

模型	训练耗时(s)	推理耗时(s)
SEN	3972.91	4.48
LPPNL	4646.87	4.89
EQT	11008.08	26.73
PhaseNet	2154.9	3.7

四个模型的训练时间和推理时间如表 4 所示,实验所采用的硬件为 NVIDIA RTX 2070 (8GB 显存)。推理时间统计为模型处理全部测试集样本(共 59904 条)的总耗时,仅计算模型运行时间,排除了数据预处理、I/O 加载等非核心环节,以尽可能消除 CPU、内存占用等外部因素对评估结果的干扰,从而更加客观地反映模型本身的计算效率。

从训练耗时来看,SEN 模型在完整训练过程中耗时为 3972.91 秒,低于 LPPNL 和 EQT,但相较于参数量最少的 PhaseNet 仍高出近一倍。虽然训练耗时较 PhaseNet 更长,但 SEN 仍然保持一个较低训练耗时的水平,并且能够获得更高的识别精度与稳定性,展现出良好的训练效率与性能折中能力。

在推理阶段,SEN 模型耗时 4.48 秒完成全部测试集的推理,优于 EQT,与 LPPNL 相当,高于 PhaseNet。虽然推理时间上还有所不足,但结合参数量以及分类和定位能力分析,SEN 可以在计算资源占用较小的情况下得到更好的分类和定位性能,且每秒推理样本数量上也基本能够满足对实时性要求较高的地震监测任务。

综上所述,本文提出的 SEN 模型在分

类性能和到时定位精度方面相对优于几个对照模型。本文模型在 P 波上具有更高的分类性能,且定位能力能够达到对照模型的水平,在 S 波上具有显著提升的分类性能,且能在维持高分类精度的同时,较好地控制到时误差的波动范围,有助于提升后续地震定位的精度和稳定性。虽然相比 PhaseNet 在轻量化上具有一定的不足,但 SEN 模型相比 LPPNL 以及 EQT 模型在参数量方面显著降低,降低了对计算资源的要求,在计算开销方面相比也具有一定程度的减少。这意味着在保证高精度的前提下,本文模型具有快速训练、快速推理以及资源占用更低的优势,可为大规模地震相检测与实时分析提供有力的技术支持。

2.4 消融实验

为验证 EMA 模块和 STFT 模块对模型性能的影响,我们设计了消融实验,对比不同模型配置下的性能表现。本次实验共设置了四个模型配置,具体如下:

基本模型:在此模型中,移除了注意力机制和 STFT 模块,仅保留基本的网络结构,用于提取时域信号的特征。

仅加入注意力模块:在基本模型的基础上,加入 EMA 注意力模块,旨在验证注意力机制对模型性能的提升。

仅加入 STFT 模块:在基本模型中,加入 STFT 模块以提取频域特征,但不使用注意力机制,评估时频信息对模型的影响。

完整模型:即本文提出的模型,同时结

合了 EMA 注意力机制和 STFT 模块，旨在充分利用时域和时频信息，提升模型性能。

在相同条件下，对以上四种模型进行了训练和测试，每个模型均训练相同次数，且使用相同的数据集进行验证。在下文中，

我们将首先对各模型在分类性能 (Precision、Recall) 上的表现进行比较，然后再分析其在定位精度 (到时误差均值与标准差) 方面的差异。

表 5 消融实验模型评价表

Table 5. Evaluation of Ablation Study Models

模型配置	Precision		Recall	
	P 波	S 波	P 波	S 波
基本模型	0.9481	0.8451	0.8418	0.6812
仅加入注意力	0.9532	0.8793	0.8689	0.7326
仅加入 STFT	0.9636	0.8821	0.8701	0.7170
完整模型	0.9579	0.8798	0.8897	0.7725

表 5 为不同模型分类性能对比情况，如表 5 所示，四个模型在 P 波识别任务上的精度都达到了 0.95 左右，表明模型对于 P 波的识别精度已经达到了较高的水平，从召回率上看也能够检测出大部分的 P 波，然而四个模型对于 S 波的表现均相对较差。基本模型在 P 波召回率指标上虽然已经取得了较高水平，但仍存在一定提升空间。在基本模型中加入注意力机制后，P 波和 S 波的召回率都有所提升。这种提升很可能是由于 EMA 注意力机制能够突出输入信号中的关键时域特征，使模型在识别关键震相特征时

更为准确和全面。基本模型中仅加 STFT 模块后，P/S 波的召回率指标也均有所提升，说明通过引入频域信息，模型可从多维度信号特征中提取出更易区分震相的模式，从而提高分类性能。完整模型在各项指标上得到了进一步的提升，说明当时域与频域特征结合注意力机制的引导，模型能够更加全面地识别震相事件，但精度相比仅加入 STFT 模块的模型而言略微下降。整体而言注意力机制与 STFT 模块的联合使用在分类性能上起到了互补的效果，使得模型识别 S 波的能力得到了显著提升。

表 6 消融实验模型到时误差表

Table 6. Arrival Time Errors of Ablation Study Models

模型配置	Mean (ms)		Std (ms)	
	P 波	S 波	P 波	S 波
基本模型	20.67	27.77	29.04	36.36
仅加入注意力	19.53	27.53	28.83	36.03
仅加入 STFT	16.09	27.39	23.53	36.24
完整模型	18.76	26.13	27.13	36.14

表 6 为不同模型下震相到时定位精度的对比情况,在使用基本模型时,P波、S波的到时误差均值与标准差相对较大,表明模型在震相到时定位方面存在较大的不确定性和误差波动。加入注意力后,震相到时的误差均值和标准差都有一定程度降低,这意味着模型在重要时序片段的聚焦能力方面较基本模型有所增强,从而在定位特征提取上更加精准,减少了整体定位误差的波动。仅加入 STFT 模块的模型也同样降低了震相到时定位误差的均值和标准差,且在 P 波的定位能力得到了显著提升,在 S 波定位方面表现与仅加入注意力的模型相当,表明仅加入 STFT 模块后模型可以从频域分辨出特定震相的频谱特征,从而更精确地捕捉震相起始点位置,使到时预测更接近真实值。在完整模型中,P波到时定位误差的均值和标准差均大于仅加入 STFT 模块的模型,说明完整模型在整合多个模块时,可能牺牲了一部分对特定模式的识别能力,转为了更泛化的发现异常,提升了对整体震相模式的建模能力和泛化能力,同时注意力机制引入了更强的通道间依赖关系,也可能在特定情形下弱化了模型对 P 波局部突变特征的直接响应,但结合分类性能来看,完整模型在 S 波的精度与召回率上具有显著提升,其在定位精度与分类稳定性之间实现了较好的权衡。在实际应用中,特别是在高噪声、复杂波形环境中,这种综合性能更为重要。

3 结论与展望

3.1 总结

本文提出了一种基于深度学习的地震震相拾取模型,该模型结合了高效多尺度注意力机制(EMA)和短时傅里叶变换(STFT),有效解决了传统方法在处理地震信号时频域信息利用不足和复杂模式下震相捕捉能力有限的问题。通过引入双分支架构,模型能够独立提取时域和频域的特征,并通过 EMA 模块对关键特征进行多尺度强化,从而提升震相识别的精度和召回率。实验结果表明,本文提出的模型在震相分类精度、召回率及定位误差等方面均表现出较好的效果,尤其在低信噪比环境下展现了较强的鲁棒性。

3.2 未来工作

尽管本研究在震相识别精度和计算效率方面取得了显著进展,但仍有进一步优化的空间。未来的工作可以从以下几个方面进行深入探索:

扩展到更复杂的观测环境:当前实验主要基于 STEAD 的公开地震数据集,这不足以应对更丰富的真实地震。Jiang (2021) 比较了 PhaseNet 与 EQT 在漾濞和玛多地震序列中的检测能力,发现这两个模型在实践中出现了不同程度的泛化性问题。因此未来在实际应用前,模型应在更大规模、更复杂观测环境产出的数据集中训练和测试模型,以验证其适用性和泛化能力。特别是在不同震源类型和复杂噪声条件下,模型的表现可

能需要进一步评估。

改进频域特征提取模块：虽然 STFT 提供了有效的时频信息，但它仍然存在窗口大小选择和频率分辨率的折中问题。未来可以探索其他时频分析方法，如小波变换方法或自适应改进 STFT 的参数设置，以进一步提升对复杂信号特征的捕捉能力。

多任务联合建模：目前的模型主要应用于震相识别方面，而在实践中模型性能往往较弱，因此可以采用集成多个模型分别监测不同物理量从而联合监测同一事件，例如蒋策等（2024）针对当前主流模型在加速度数据上性能较弱的问题，基于 PhaseNet 提出了两个分别适配速度与加速度记录的模型，并构建了完整的地震自动处理流程，在新丰江水库实际地震中取得了 3.8 倍于人工目录的检测数量。未来还可以尝试将震相拾取与其他地震学任务（如震源定位、震级估计等）进行联合建模，提升模型的综合性能。这样不仅能够提高地震事件检测的准确性，还可以为震后分析和应急响应提供更全面的支撑。

模型部署与实时应用：为了使模型适应实际的地震监测需求，未来还需进一步优化模型的推理速度和计算资源消耗，确保其在实时数据流中的快速响应能力。同时，针对边缘计算平台的部署，还可以通过模型压缩和加速技术，进一步降低计算开销，提高模型的实际应用价值。

总的来说，本文提出的 SEN 模型在地

震信号自动化识别与处理方面为地震学研究提供了新的思路和技术支持。随着更多数据的积累和算法的不断优化，未来该模型有望在实际应用尤其在实时地震监测和早期预警系统中发挥作用。

References

- Allen, J.B., 1977. Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(3), 235 - 238. doi:10.1109/TASSP.1977.1162950.
- Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine Learning for Data-Driven Discovery in Solid Earth Geoscience. *Science*, 363(6433), eaau0323. doi:10.1126/science.aau0323.
- Chai, C., 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. doi:10.1029/2020GL088651.
- Chen GuoYi, Yang Wen, Tan YuYang, Zhang HaiJiang, LI JunLun. 2023. Automatic phase detection and arrival picking for micro-seismic events in hydraulic fracturing based on machine learning and array correlation. *Chinese Journal of Geophysics* (in Chinese), 66(4): 1558-1574, doi:

- 10.6038/cjg2022P0542.
- Chen, Y., Zhang, G., Bai, M., et al., 2019. Automatic Waveform Classification and Arrival Picking Based on Convolutional Neural Network. *Earth and Space Science*, 6(7), 1244 - 1261. doi:10.1029/2018EA000466.
- Hu Jinjun, Ding Yitian, Zhang Hui, Jin Chaoyue, Tang Chao, 2023. A Real-Time Seismic Intensity Prediction Model Based on Long Short-Term Memory Neural Network. *Earth Science(in Chinese)*, 48(5): 1853-1864. doi: 10.3799/dqkx.2022.338.
- He Bin, Zhou Yunyao, Lü Yongqing. 2024. Seismic First Break Picking Algorithm Combining U-net and FPN [J]. *Journal of Geomatics(in Chinese)*, 49 (1) : 82-87. 10. doi:14188/j. 2095-6045. 2022180.
- Jiang, C., Fang, L., Fan, L., 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. doi:10.29382/eqs-2021-0038.
- Jiang Ce, Lü Zuoyong, Fang Lihua, 2024. Earthquake Detection Model Trained on Velocity and Acceleration Records and Its Application in Xinfengjiang Reservoir. *Earth Science(in Chinese)*, 49(2): 469-479. doi: 10.3799/dqkx.2023.186.
- Lan, B., Zhao, S., Zeng, H., et al., 2025. Seismic Phase Picking Using a Cross-Attention Network on NVIDIA Jetson Xavier NX. *IEEE Access*, 12. doi:10.1109/ACCESS.2024.3471848.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. *Nature*, 521(7553), 436 - 444. doi:10.1038/nature14539.
- Li, B., Fan, L., Jiang, C., Liao, S., Fang, L., 2023. CSESnet: A Deep Learning P-Wave Detection Model Based on UNet++ Designed for China Seismic Experimental Site. *Frontiers in Earth Science*, 10, 1032839. doi:10.3389/feart.2022.1032839.
- Li, H., Li, J., Li, X., et al., 2025. Seismic Picking Attention Module. *IEEE Transactions on Geoscience and Remote Sensing*, PP. doi:10.1109/TGRS.2024.3476329.
- Li, W., Chakraborty, M., Fenner, D., et al., 2021. EPick: Multi-Class Attention-based U-shaped Neural Network for Earthquake Detection and Seismic Phase Picking. *arXiv*. doi:10.48550/arXiv.2109.02567.
- Liao, W.Y., Lee, E.J., Mu, D., et al., 2021. ARRU Phase Picker: Attention Recurrent-Residual U-Net for Picking Seismic P- and S-Phase Arrivals. *Seismological Research Letters*, 92(4), 2293 - 2303. doi:10.1785/0220200382.
- Lin, T.Y., Goyal, P., Girshick, R., et al.,

2017. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318 - 327. doi:10.1109/TPAMI.2018.2858826.
- Mousavi, S.M., Beroza, G.C., 2022. Deep-Learning Seismology. *Science*, 377, eabm4470. doi:10.1126/science.abm4470.
- Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y., Beroza, G.C., 2020. Earthquake Transformer—An Attentive Deep-Learning Model for Simultaneous Earthquake Detection and Phase Picking. *Nature Communications*, 11, 3952. doi:10.1038/s41467-020-17591-w.
- Mousavi, S.M., Sheng, Y., Zhu, W., et al., 2019. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7, 179464 - 179476. doi:10.1109/ACCESS.2019.2947848.
- Ouyang, D., He, S., Zhan, J., et al., 2023. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *arXiv*, abs/2305.13563. doi:10.1109/ICASSP49357.2023.10096516.
- Ross, Z.E., Meier, M.A., Hauksson, E., 2018. Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, 108(5A), 2894 - 2901. doi:10.1785/0120180080.
- Saad, O.M., Chen, Y., 2021. CapsPhase: Capsule Neural Network for Seismic Phase Classification and Picking. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1 - 11. doi:10.1109/TGRS.2021.3089929.
- Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention Is All You Need. *arXiv*. doi:10.48550/arXiv.1706.03762.
- Wang, J., Xiao, Z., Liu, C., Zhao, D., Yao, Z., 2019. Deep-Learning for Picking Seismic Arrival Times. *Journal of Geophysical Research: Solid Earth*, 124, 6612 - 6624. doi:10.1029/2019JB017536
- 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. doi:10.1029/2020JB021444.
- Ye, H., Chen, J., Gong, S., et al., 2024. ATFNet: Adaptive Time-Frequency Ensembled Network for Long-term Time Series Forecasting. *arXiv preprint*, arXiv:2404.05192. doi:10.48550/arXiv.2404.05192.
- Yu, Z., Wang, W., 2022. LPPN: A Lightweight Network for Fast Phase Picking. *Seismological Research Letters*, 93(5), 2834 - 2846. doi:10.1785/0220210309.

- Yu, Z., Wang, W., Chen, Y., 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:10.1016/j. eqs.2022.10.001.
- Zhang, J., Li, Z., Zhang, J., 2023. Simultaneous Seismic Phase Picking and Polarity Determination with an Attention-Based Neural Network. *Seismological Research Letters*. doi:10.1785/0220220247.
- Zhou, T., Ma, Z., Wen, Q., et al., 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. *arXiv e-prints*. doi:10.48550/arXiv.2201.12740.
- Zhou, Y., Yue, H., Kong, Q., 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.
- Zhu, L., et al., 2019. Deep Learning for Seismic Phase Detection and Picking in the Aftershock Zone of 2008 Mw7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, 293, 106261. doi:10.1016/j. pepi.2019.05.004.
- Zhu, W., Beroza, G. C., 2019. 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., Biondi, E., Li, J., et al., 2023. Seismic Arrival-Time Picking on Distributed Acoustic Sensing Data Using Semi-Supervised Learning. *Nature Communications*, 14(1), 8192. doi:10.1038/s41467-023-43355-3.
- 附中文参考文献
- 陈国艺, 杨文, 谭玉阳等. 2023. 基于机器学习和台阵相关性的水力压裂微地震事件自动识别及到时拾取. *地球物理学报*, 66(4):1558-1574, doi:10.6038/cig2022P0542.
- 胡进军, 丁祎天, 张辉, 靳超越, 汤超, 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. doi:10.3799/dqkx.2022.338.
- 何彬, 周云耀, 吕永清. 2024. 结合 U-net 与 FPN 的地震初至波拾取算法[J]. *测绘地理信息*, 49(1). doi:14188/j. 2095-6045. 2022180.
- 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[J]. *地球科学*, 49(2): 469-479. doi:10.3799/dqkx. 2023.186.