Toward an AI-Enhanced Near-Real-Time Earthquake Cataloging System for the Southern California Seismic Network (SCSN)

Yijian Zhou, Ryan Tam, Gabrielle Tepp, Rayomand Bhadha, Ellen Yu, Zach Ross, Zhongwen Zhan, & Allen Husker
GPS & SeismoLab, California Institute of Technology

Email: zhouyj@caltech.edu

Abstract

Rapid and complete microseismicity monitoring is increasingly important as modern seismic networks evolve, supporting scientific investigations, hazard assessment, and timely situational awareness following significant earthquakes. Although recent AI/ML algorithms have substantially improved phase picking, their integration into routine regional network operations has remained limited. Here, we present the development of an AI-enhanced near-real-time cataloging framework for the Southern California Seismic Network (SCSN).

This framework separates offline model training from online inference. Continuous waveform data are processed using modular AI phase pickers, followed by subnetwork-based phase association and optional refinement of arrival times using pickers pre-trained on a California-wide positive dataset. The resulting phase files are subsequently processed by the existing SCSN location and magnitude modules, ensuring compatibility with current operational products while allowing AI components to evolve independently. The modular design also enables straightforward integration and comparison of multiple phase picking and association algorithms.

The offline workflow constructs an enhanced reference catalog for 2020–2025 by processing continuous waveform data archived in the SCEDC AWS Dataset. We first apply a rule-based workflow Phase, Association, and Location (PAL; Zhou et al. 2021) to generate positive and negative training samples for individual station-day waveform segments, which are used to train four AI pickers: Self-Attention RNN (SAR), PhaseNet, Res-UNet, and Frame-Transformer. These SCSN-optimized pickers are deployed in the online workflow while simultaneously producing an enhanced retrospective catalog for 2020–2025.

The online workflow processes continuously incoming waveform data using overlapping 10-minute windows with 5-min steps. Within each window, every picker is applied using a short 2.5-s sliding stride, so that the same seismic arrival is evaluated multiple times at different 25-s window positions. Candidate arrivals are retained when they are detected repeatedly by the same model across overlapping windows and/or by multiple pickers. Phase association is then performed independently within overlapped subnetworks, after which detections are merged to remove duplicates. This modular strategy provides a pathway for incorporating future advances in earthquake detection and phase association into operational cataloging.

1. SCSN & Division of Sub-networks

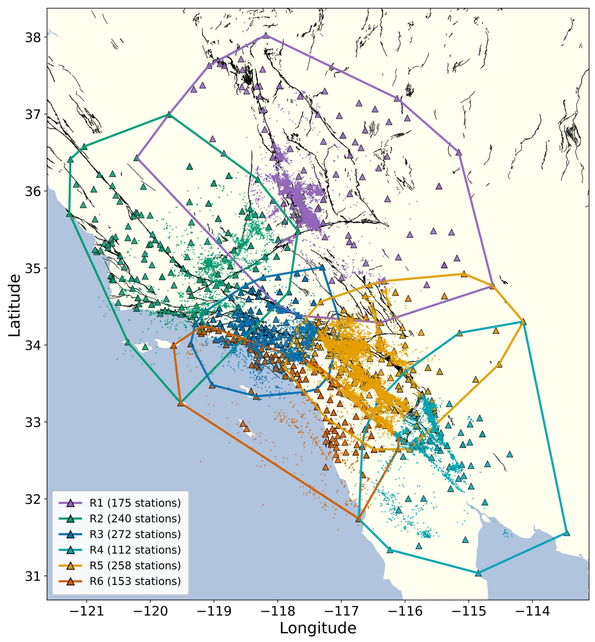
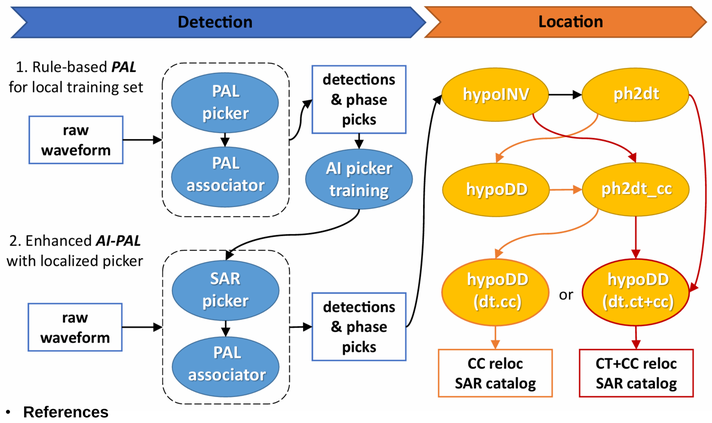


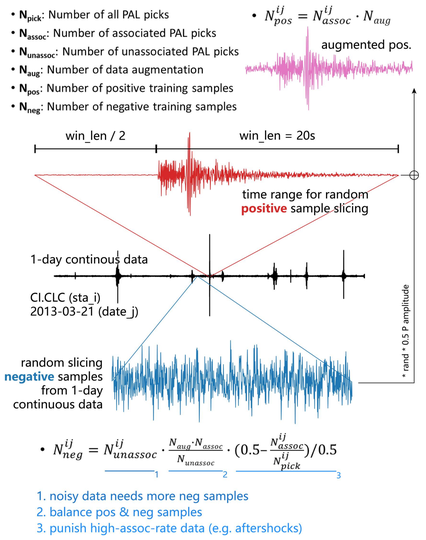
Figure 1. Division of sub-networks for SCSN. Colored dots show the seismicity clusters used to define each region (relocated SCSN catalog, 1981-2024), triangles mark the selected stations, and colored outlines enclose the final station coverage for each subnetwork..

2. Upgraded AI-PAL Workflow

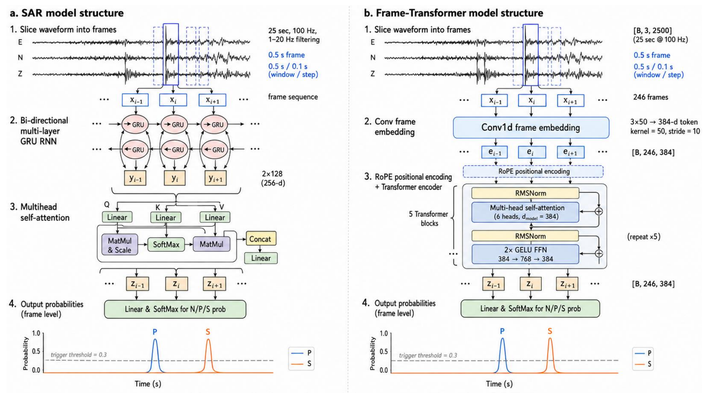


- **References**

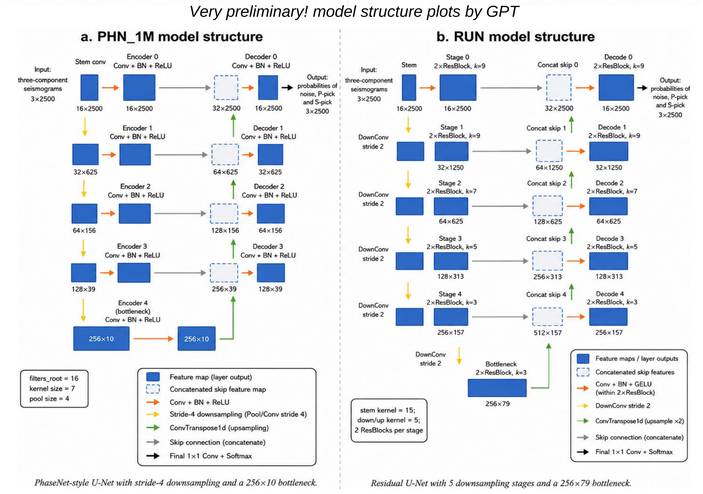
- **Zhou et al., JGR-SE (2025)**, AI-PAL: Self-Supervised AI Phase Picking via Rule-based Algorithm for Generalized Earthquake Detection, https://doi.org/10.1029/2025JB031294
- **Yijian ZHOU. (2026).** YijianZhou/AI-PAL: AI-PAL (Version v7.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.22054970



Multi-picker Ensemble: Model Type & Training Data



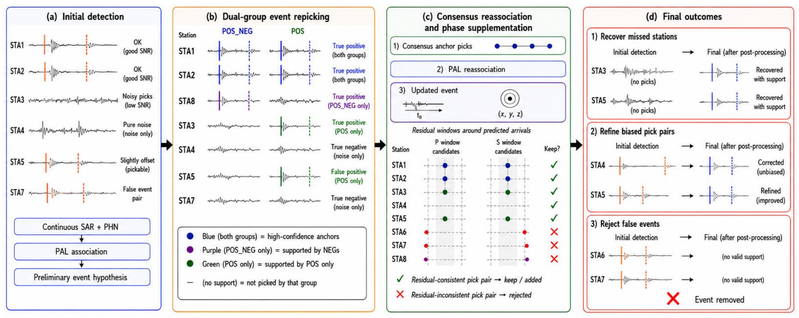
Very preliminary! model structure plots by GPT



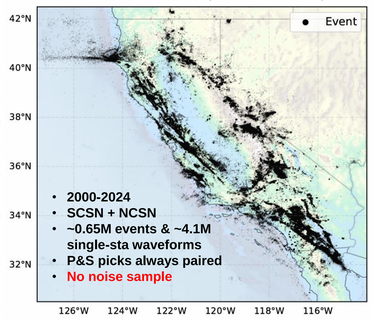
PhaseNet-style U-Net with stride-4 downsampling and a 256×10 bottleneck.

Residual U-Net with 5 downsampling stages and a 256×79 bottleneck.

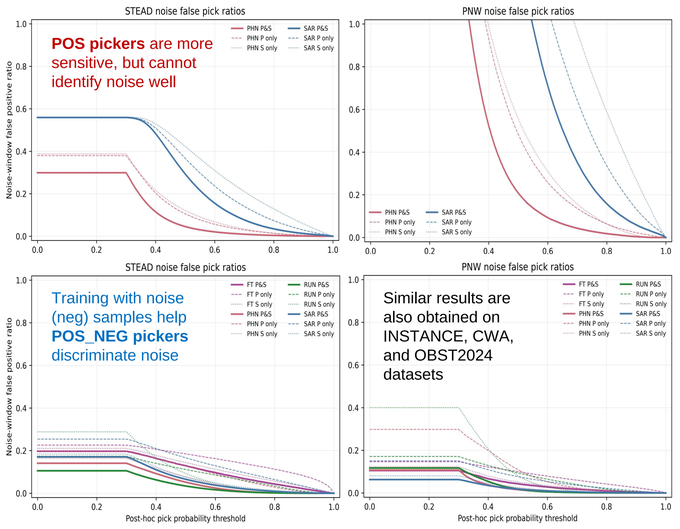
Event-level post-processing in AI-PAL V7 *(very preliminary schematic plot by GPT)*



CEED Pos Dataset (Zhu et al., 2026)



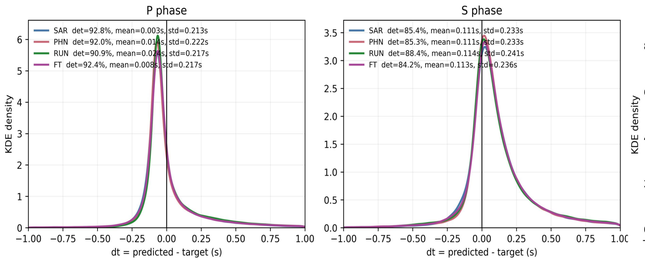
3. Picker Benchmark on SeisBench Datasets



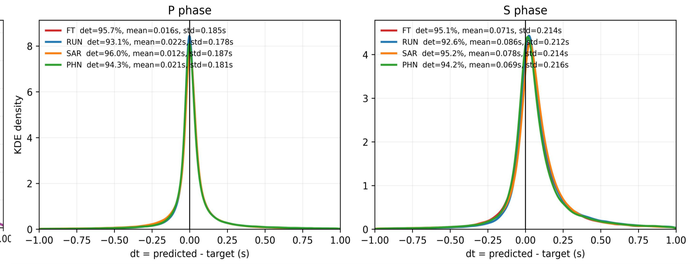
Conclusions

- POS_NEG pickers yield slightly lower detection completeness, but much higher accuracy for noise samples, thus is suitable for continuous data processing
- POS pickers should be applied to positive samples, i.e. for repicking, after an initial detection is made, so as to recover missed weak phases

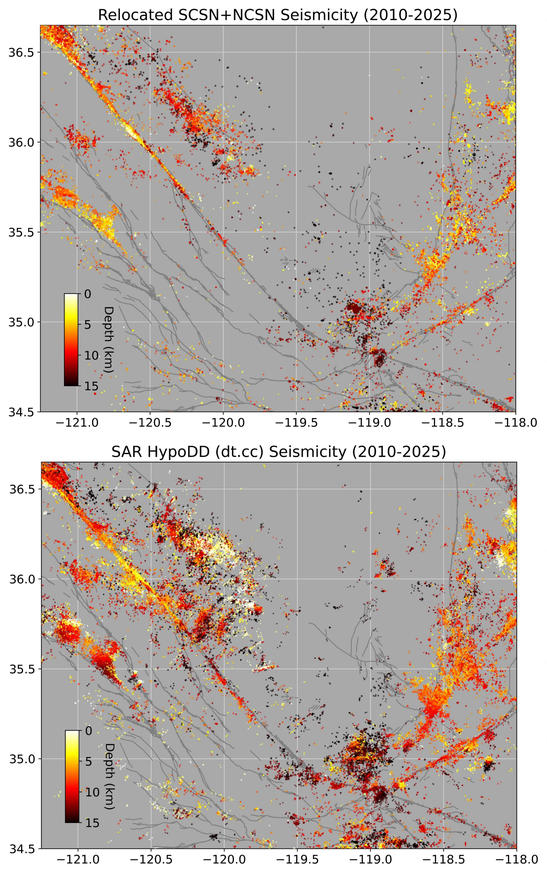
Cent_Cal POS_NEG picker on INSTANCE positive samples



CEED POS picker on INSTANCE positive samples



4. Preliminary results in Central California and the SoCal Real-time Data flow



(This is based on previous version of AI-PAL that adopts single-picker & no post-processing)