

Article

Symmetry-Aware Collaborative Attention Network for Robust Weak Seismic Phase Picking

Yunpeng Wang ¹, Qing Li ², Chao Zhang ³, Yatong Bai ¹, Xiaofei Du ⁴, Jianfeng Wang ^{1,*} and Yuda He ^{5,*}

¹ School of Software, Taiyuan University of Technology, Taiyuan 030024, China; 2023005944@link.tyut.edu.cn (Y.W.)

² Simulation Training Center, Officers College of PAP, Chengdu 610000, China; liqing0821@outlook.com

³ School of Intelligent Equipment Manufacturing, Guangxi Vocational & Technical Institute of Industry, Nanning 530001, China; zhangchao@gxgy.edu.cn

⁴ School of Mechanical Engineering, Nanjing Institute of Technology, Nanjing 211167, China

⁵ Kuala Lumpur Campus, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia

* Correspondence: wangjianfeng@tyut.edu.cn (J.W.); heyuda@graduate.utm.my (Y.H.)

Abstract

Reliable seismic phase picking is essential to earthquake monitoring, as it fundamentally affects earthquake location and source inversion. In challenging field conditions, non-stationary waveforms, diverse morphological features and intense background noise all hinder the detection of weak phases. Seismic time series also exhibit inherent spatiotemporal asymmetry. Nevertheless, mainstream networks rely on symmetric and uniform feature extraction strategies. They overlook asymmetric properties of waveforms and introduce additional picking errors. We therefore present SymPhase, a symmetry-aware collaborative attention network, to achieve precise and robust P- and S-phase picking. Using a 1D encoder–decoder backbone, the model combines global enhancement and local refinement. It captures both long-range dependencies and local features, reducing missed weak-phase detections and minimizing arrival-time bias. Extensive tests are conducted on the CEED and DiTing datasets. The results demonstrate that SymPhase outperforms both TCN and PhaseNet. On the CEED dataset, the F1 scores for P and S phases are 0.9797 and 0.9006, with mean absolute errors of 0.0761 s and 0.1003 s. On the difficult DiTing dataset, the S-phase F1 score reaches 0.4724 with a corresponding error of 0.7386 s. These results validate its superior performance for weak signal recognition. With strong accuracy and noise robustness, SymPhase provides a viable solution for automated earthquake monitoring systems.

Keywords: seismic phase picking; SymPhase; symmetry-aware attention; collaborative attention; weak-phase detection; 1D encoder–decoder; deep learning



Academic Editor: Jie Yang

Received: 4 June 2026

Revised: 4 July 2026

Accepted: 9 July 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

P- and S-wave arrival picking forms the fundamental foundation of seismic waveform interpretation and earthquake monitoring workflows [1]. In our practical data processing work, accurate pick results are indispensable for earthquake relocation calculations and source parameter inversion; these timestamp benchmarks also supply key boundary constraints when building seismic catalogs and exploring regional tectonic characteristics [2]. From years of sorting published seismic research, it is easy to find that the reliability of all follow-up seismological analyses hinges heavily on the quality of original phase identification.

Classic empirical picking algorithms such as STA/LTA and AIC depend entirely on artificially set threshold parameters and empirical judgment [2–4]. When tuning these algorithms on measured field waveforms, we often notice a prominent flaw: their performance degrades rapidly once ambient noise intensifies and SNR falls, making them hard to generalize to varied field geological conditions. The rise of deep learning effectively breaks this limitation. As a pioneering U-Net-based picking scheme, PhaseNet calculates the probability distribution of P and S arrivals directly from unprocessed original seismic traces and pioneers data-automatic picking ideas [5]. Later studies embedded CNN, RNN and self-attention modules to boost feature extraction capacity, which gave birth to widely adopted models like EQTransformer and EPick [6,7]. Recent studies lean toward multi-scale fusion and CNN–Transformer hybrid designs, which balance local waveform mutation and long-time-series correlation and deliver stable performance under conventional monitoring datasets [8–11].

More broadly, data-driven and learning-based methods have advanced intelligent decision-making in aerospace guidance and cooperative trajectory planning [12,13], IoT-enabled secure mobility prediction [14], data-driven engineering design and neural-network applications [15–17], and intelligent computing together with zero-trust security and emerging AI governance frameworks [18–20]; building on this broader trend, similar deep-learning paradigms have been increasingly adopted in seismology for robust automated phase picking [5–7].

Despite these successive structural improvements across existing algorithms, one core physical property of seismic waveforms has long been overlooked by most designs. During repeated dataset validation on CEED and DiTing, we found nearly all mainstream networks adopt unified symmetric convolution rules across full input sequences, ignoring intrinsic spatiotemporal asymmetry inside real seismic waveforms. Such fixed modeling logic fails to match varied waveform characteristics between different seismic phases. P-wave signals pop up with sharp, short-duration onset, whereas S-wave energy accumulates slowly and forms loose, wide waveform envelopes. Noise interference will further widen such asymmetric differences. This inconsistency between symmetric network design and asymmetric seismic properties inevitably damages model feature extraction capacity, leaving plenty of weak-phase events unpicked and bringing unpredictable timing errors.

Apart from the above physical-model mismatch, three concrete technical defects still hinder high-precision picking amid low-SNR surroundings. First, seismic traces feature prominent nonstationarity and uneven time-frequency distribution. Current multi-scale modules mostly expand receptive fields empirically without targeted distinction on P- and S-wave morphological differences, which explains their poor performance facing faint S-wave signals. Second, most attention modules deployed in seismic prediction lack customized optimization toward seismic waveform evolution rules; single-type attention struggles to capture long-distance waveform correlation under heavy ambient noise. Third, vanilla skip connections inherited from original U-Net transmit all shallow low-level features without screening, dragging massive useless noise and redundant feature information into decoder branches. Gradually accumulated clutter blurs phase boundary features and becomes a major cause of missed weak-phase identification.

To fix the aforementioned drawbacks, we put forward SymPhase, a symmetry-aware collaborative attention network targeting precise P/S picking under noisy complicated field environments. The network builds upon a 1D encoder–decoder skeleton optimized for nonstationary seismic time series and assembles three core components: global channel-time attention, optimized gated skip links and multi-scale feature extraction branches. In our network construction, we adopt 1D CBAM to extract global contextual information across entire seismic traces [21]. Different from ordinary skip link designs, our gated

attention relies on high-level semantic features output by the decoder to screen useless noise from low-level encoder features [22]. We set three groups of convolution kernels for multi-scale branches purposely, to adapt the transient spike of P onset and the slowly varying energy envelope of S wave respectively [7]. From our ablation test results, this coordinated design of global restriction plus local screening evidently reduces the leakage of weak-phase signals and lowers arrival offset errors.

We summarize four core innovations of this study below:

- **Symmetric-asymmetric collaborative modeling framework.** We tailor SymPhase exclusively for seismic phase picking. Its dual-path feature learning combines globally symmetric contextual restriction and locally asymmetric adaptive weight adjustment. This innovative structure effectively bridges the gap between traditional unified network modeling and naturally asymmetric seismic waveforms.
- **Global-local dual attention integration.** A standalone 1D CBAM unit strengthens feature representation from channel and temporal dimensions globally, meanwhile gated skip attention selectively sifts valid local features in skip paths. Paired usage of the two attention designs balances overall semantic consistency and local detail capture and lifts model performance heavily in low-SNR scenarios.
- **Multi-scale adaptive feature extraction.** Multi-branch convolution with different kernel sizes is deployed to adapt to divergent temporal scales of P and S waveforms, strengthening model adaptability toward faint phase signals, complex background noise and anomalous seismic recordings.
- **Comprehensive experimental validation.** We carry out abundant ablation tests and cross-model comparison experiments to verify the effectiveness of every proposed module. Measured data across standard and difficult datasets confirm SymPhase beats existing mainstream algorithms in picking accuracy, stability and anti-noise capability.

2. Related Work

2.1. Seismic Phase Picking Methods

Seismic waveforms are modulated by source properties, propagation routes and near-surface geology together, plus ever-changing ambient noise, which makes recorded traces highly nonstationary. In our dataset sorting work, low-SNR cases commonly obscure real P/S onsets and trigger missing or false picks [1,3,4]. Traditional STA/LTA-, AIC- and AR-based algorithms judge arrival moments relying on energy fluctuation or statistical rules [2]. These methods are physically intuitive and easy to deploy, yet their performance relies heavily on manually tuned thresholds, so they struggle to cope with complicated field noise and limited robustness is a frequent drawback. To overcome such drawbacks, data-oriented deep learning gradually became a viable alternative to extract inherent features directly from raw seismic recordings.

The core merit of deep learning lies in escaping manual threshold setup. Originally, ASPP was designed for dense segmentation in computer vision [23], and researchers later transplanted similar multi-scale ideas into seismic picking. Early CNN attempts verified convolutional networks' practicability for automatic phase labeling [24]. As a milestone, PhaseNet outputs independent probability curves for P, S and noise components and established a solid CNN baseline in this field [5]. EQTransformer first introduced an attention mechanism into joint detection and picking tasks, opening a new path toward global dependency learning [6]. Follow-up works adopted Transformer and unsupervised algorithms to lift microseismic weak-phase accuracy; TCN and Inception multi-branch structures were also introduced to fit long sequences and multi-level feature fusion [10,11,25–28]. Recent CNN–Transformer hybrid and multi-scale networks further expand this technical line [8,9].

Several studies have further explored one-dimensional attention mechanisms tailored to seismic waveforms. EPick embeds attention blocks in a multi-scale U-Net to emphasize informative encoder features during 1D segmentation-style picking [7]. ARRU integrates soft attention gates and recurrent-residual units into a U-Net backbone to suppress irrelevant waveform responses and strengthen temporal connections in 1D seismograms [29]. Semi-channel attention and CNN–Transformer hybrids have been proposed to recalibrate multi-scale 1D features for phase picking [8,9], and channel-attention modules combined with convolution-enhanced attention have been applied to earthquake detection and picking on 1D traces [30]. These approaches demonstrate growing interest in attention-based refinement of 1D seismic features; however, most existing designs rely on a single attention route or treat global and local attention separately. SymPhase addresses this gap by coordinating global 1D CBAM recalibration with attention-gated skip selection within one encoder–decoder framework.

Aside from network innovations, publicly available datasets make cross-model benchmarking feasible. A series of tests based on global seismic network data show accuracy and inference cost always conflict among different neural pickers [31], a trend we also observed during our model comparison. Recent review articles have summarized field advances and unresolved challenges from dataset and algorithm dimensions [32].

Aiming at practical station deployment demands, scholars have developed compact regional networks such as LPPN and CSESnet to adapt local observational environments [33,34].

Despite abundant previous improvements, few existing algorithms specially curb specific errors like frequent S-wave missing and fuzzy weak onset under heavy noise. Inspired by the above research gaps, we build our model via multi-scale stacking, global semantic enhancement and adaptive skip feature filtering to remedy this deficiency.

2.2. Multi-Scale Feature Extraction in Seismic Picking

We can easily observe distinct multi-scale traits across seismic waveforms: P-wave onset emerges as a narrow, sharp energy spike, whereas S-wave energy evolves slowly to form a wide-range envelope. Changes in epicenter–station distance, propagation medium and site geology further widen such scale divergence [7,8]. Single-size convolution possesses a fixed receptive field size, which cannot simultaneously capture transient abrupt onset and slow-varying long-period waveform information.

To tackle this problem, existing research commonly arranges parallel convolutions with diverse kernel dimensions or builds layered encoder–decoder structures to collect local fine details and global context separately [7,8,35]. Small-kernel branches lock onto abrupt local waveform mutation, while large-scale branches trace long-term energy variation and background noise distribution. Feature aggregation combines merits from both ends to depict complicated phase distribution rules precisely.

Still, mainstream multi-scale design stays confined to receptive field expansion and simple feature fusion. Few designs specially differentiate morphological discrepancy between P and S phases or impose targeted constraints for effective phase feature learning. Within SymPhase, multi-scale convolution works as the base feature extraction module. It first distinguishes intrinsic P/S waveform differences and feeds abundant multi-dimension features into subsequent CBAM global attention and gated local filtering modules, forming coordinated optimization to break picking bottlenecks under noisy environments.

2.3. Global Context Enhancement

In our early trials using plain convolution or single-dimensional attention modules, we found such structures cannot balance global waveform constraints and local phase

positioning well amid heavy noise. Biased feature optimization tends to erase subtle weak-phase features or overemphasize irrelevant noise components [5,7].

We adopt dual-dimension feature recalibration to tackle this drawback. This method conducts progressive feature optimization to coordinate global semantics and local detail expression [21]. Channel pooling aggregates holistic waveform information first, while temporal focusing locks down possible phase occurrence ranges. We tailor this mechanism to fit 1D seismic sequences via targeted adaptation for our network construction.

Unfortunately, this global optimization only improves holistic feature expression and lacks noise elimination on skip connection branches, making it hard to refine subtle phase onset details locally. We configure the above strategy as our global enhancement branch in SymPhase. It stabilizes long-span waveform semantic information and complements the local feature selection unit to realize joint global–local optimization and raise final picking precision.

2.4. Local Feature Selection

Skip connection in encoder–decoder retains shallow fine-grained features, yet unfiltered raw information inevitably mixes messy noise, which severely disturbs positioning of faint P and S onset under complex ambient surroundings [22].

Semantic-guided screening introduces high-level abstract features as a screening benchmark to assign adaptive weight to shallow skip features, reserving only waveform components closely related to real seismic phases [22]. High-layer semantic information owns stronger classification capacity, guiding the network to concentrate on valid phase-related texture and suppress irrelevant noise disturbance.

The current literature mostly applies such local filtering as a trivial auxiliary for the skip connection upgrade, ignoring its huge potential on fine-grained seismic phase picking. Different from global overall adjustment, local selection targets tiny waveform variation and blocks redundant noisy feature transmission effectively. SymPhase embeds this mechanism into the independent local optimization branch, constructing a complete “global enhancement plus local screening” collaborative architecture for reliable waveform modeling and detail extraction in tough picking scenarios.

2.5. Symmetry and Asymmetry in Time-Series Modeling

From our summary of existing picking frameworks, most mainstream networks borrow uniform symmetric convolution rules from conventional time-series algorithms and apply identical transformation over all temporal positions and feature channels. This preset symmetric distribution hypothesis does not match real seismic traits, since abrupt P onset and slowly evolving S envelopes create inherent temporal asymmetry. In feature extraction computation, such inconsistency easily masks asymmetric waveform details and yields blurry prediction peaks, leading to frequent loss of weak phase points.

Existing research published in Symmetry demonstrates that Li et al. took waveform symmetry as critical indicator for microseismic first-arrival identification and validated its anti-noise capability [36]. Drawing on their research idea, we extend such symmetry-related modeling from microseismic scenarios to conventional P and S picking and fully exploit intrinsic morphological differences of real seismic traces.

To fix the aforementioned mismatch problem, our framework combines two complementary modeling strategies. We keep globally consistent waveform constraints via symmetric context construction and adopt locally adjustable weight parameters to implement phase-oriented optimization for balanced global semantics and focused local features.

A quantitative evaluation of waveform morphological asymmetry and onset SNR distributions on CEED and DiTing is reported in Section 4.1.

jointly raising identification reliability under noisy field data. Following our proposed symmetric–asymmetric design idea, the network applies unified global constraint plus locally adjustable weight to fit the inherent waveform asymmetry of real seismic recordings.

3.2. Backbone Network

SymPhase adopts five-layer 1D U-Net as the baseline structure, with three input channels matching three-component observation and three output categories. Every encoder block combines multi-scale convolution and strided downsampling to enlarge the receptive field and compress high-level features. The decoder uses transposed convolution for upsampling, and all skip paths complete temporal alignment before feature concatenation. Merged features go through extra multi-scale convolution to recover original sample length. On the basis of mature PhaseNet architecture, we embed multi-scale block and dual-attention units to improve adaptability to complicated measured waveforms.

3.3. Multi-Scale Convolution Module

P and S components differ obviously in temporal characteristics: P emerges as an instantaneous sharp spike while S evolves into a wide continuous energy envelope, making single-kernel convolution unable to capture both features simultaneously. To solve this problem, every basic block is equipped with three parallel convolution branches targeting distinct time ranges:

- **Small-scale branch:** Focuses on abrupt P-wave starting points.
- **Medium-scale branch:** Extracts intermediate waveform transition features.
- **Large-scale branch:** Tracks long-term energy change of S-wave envelopes.

Let the outputs of the three branches be (F_1, F_2, F_3) . The multi-scale features are concatenated along the channel dimension and fused via convolution, batch normalization, and ReLU activation.

$$F = \phi([F_1; F_2; F_3]),$$

where $[\cdot; \cdot; \cdot]$ denotes channel-wise concatenation and $\phi(\cdot)$ represents the feature fusion operation. The fused multi-scale features serve abundant alternative information for subsequent global refinement and local screening, laying a solid foundation for accurate phase identification.

3.4. Global Context Enhancement Branch

SymPhase adopts a 1D variant of the Convolutional Block Attention Module (CBAM) [21]. The original CBAM operates on 2D feature maps $F \in \mathbb{R}^{B \times C \times H \times W}$ and applies channel attention followed by spatial attention. For 1D seismic features $F \in \mathbb{R}^{B \times C \times T}$, the spatial axes (H, W) are replaced by the temporal axis T : channel attention uses global average and max pooling along time, and the spatial attention branch is reinterpreted as *temporal attention* by aggregating channel-wise statistics at each time sample and convolving along the sequence with a 1D kernel ($k = 7$). The sequential attention order of CBAM is preserved.

Compared with the standard 2D CBAM, the implemented module differs only in dimensionality and operator type: 2D pooling and Conv2d are replaced by 1D pooling and Conv1d, and the channel MLP is implemented as two 1×1 Conv1d layers, which is the conventional 1D equivalent. SymPhase first extracts multi-scale features through parallel Conv1d branches with kernel sizes 3, 7, and 15, fuses them, and then applies this 1D CBAM to the fused feature map for global recalibration. The temporal attention kernel $k = 7$ follows the original 2D CBAM spatial branch and is independent of the multi-scale kernel set. The module works together with gated skip attention for local refinement.

The channel–time recalibration comprises two sub-modules:

- **Channel weight optimization:**

$$M_c(F) = \sigma(\text{FC}(\text{AvgPool}(F)) + \text{FC}(\text{MaxPool}(F))), \quad F' = M_c(F) \odot F.$$

- **Temporal attention:**

$$M_t(F') = \sigma(\text{Conv}([\text{AvgPool}_c(F'); \text{MaxPool}_c(F')])), \quad F'' = M_t(F') \odot F'.$$

Here, $\sigma(\cdot)$ denotes the sigmoid function, and $\odot$ is element-wise multiplication. This branch emphasizes global semantic features, providing stable high-level representations for the decoder, but does not perform fine-grained local selection, complementing the local feature selection branch.

3.5. Local Feature Selection Branch

The gated skip connection in SymPhase is adapted from the attention gate proposed in Attention U-Net [22]. In the original 2D formulation, an encoder skip feature and a decoder gating signal are linearly projected, combined, and passed through a sigmoid to obtain a spatial attention map that reweights the skip feature before fusion. For 1D seismic sequences, we implement the same design along the temporal axis using 1×1 Conv1d projections. Let $x_{\text{skip}} \in \mathbb{R}^{B \times C_x \times T}$ denote the encoder skip feature and $g \in \mathbb{R}^{B \times C_g \times T}$ the coarser-scale decoder feature used as the gating signal. The attention gate is defined as

$$q_x = W_x * x_{\text{skip}}, \quad q_g = W_g * g, \\ \alpha = \sigma(W_\psi * \text{ReLU}(q_x + q_g)), \quad \tilde{x}_{\text{skip}} = \alpha \odot x_{\text{skip}},$$

where $*$ denotes 1×1 Conv1d, $\sigma(\cdot)$ is the sigmoid function, and $\odot$ is element-wise multiplication. The operators W_x , W_g , and W_ψ correspond to the skip, gating, and attention-output projections in the original Attention U-Net [22], here implemented as learnable 1D convolutions. The resulting coefficient $\alpha \in \mathbb{R}^{B \times 1 \times T}$ acts as a temporal attention map. The filtered feature $\tilde{x}_{\text{skip}}$ is concatenated with upsampled decoder features for subsequent fusion. In SymPhase, g carries high-level decoder semantics that suppress irrelevant shallow noise in x_{skip} while preserving phase-related transients; this local gate complements the global CBAM branch in Section 3.4, which recalibrates fused multi-scale features over the full sequence.

Although skip connections preserve shallow feature details, they can carry noise and redundant information, which severely degrades local arrival localization accuracy, especially for weak P/S onsets under high noise levels. SymPhase therefore applies the above gate as a semantic-guided local feature selection on each skip path.

3.6. Inference and Post-Processing

We screen suitable hyperparameters via grid search on validation subsets, with F1 score set as the core screening indicator to strike a balance between precision and recall. From repeated tuning trials, appropriate parameter adjustment can marginally sacrifice precision yet raise overall recall and F1 values, which effectively stabilizes the identification performance of faint seismic phases.

We adopt separate cutoff values for P and S to eliminate evaluation bias, namely the dual-threshold scheme. All candidate thresholds are picked from the range $\{0.10, 0.15, \dots, 0.95\}$ through traversal on validation data, and the pair $\text{thr}_P, \text{thr}_S$ that maximizes F1 will be fixed for all subsequent test evaluations.

Predictions count as correct picks when their offset against real arrival points stays within ten sampling points, equivalent to 0.1 s under the 100 Hz sampling rate. Under this judging rule, we calculate precision, recall and F1 for two wave types, alongside residual mean, standard deviation and MAE, to comprehensively assess classification capacity and temporal positioning error.

The above post-processing pipeline delivers steady picking outcomes under noisy, low-SNR waveform conditions, which creates credible evaluation standards for ablation tests and cross-model comparison.

4. Experiments

4.1. Datasets and Preprocessing

All model tests in this work rely on two open-access seismic datasets: CEED and DiTing [37,38]. These two datasets have become mainstream benchmark resources for phase picking algorithm validation, and all samples contain three-component seismic waveforms alongside manually calibrated P- and S-wave arrival annotations. To guarantee reproducible data division, we locked random seeds in advance before segmentation and divided the whole dataset into training, validation and test groups with a common 7:2:1 split ratio. After sorting and inspecting waveform samples repeatedly, we found CEED possesses relatively neat waveform quality, moderate signal-to-noise ratios and balanced magnitude distribution across contained events, hence we take it as the primary resource for model training and ablation tests of each functional module.

Different from CEED's curated samples, DiTing is a large-scale seismic dataset collected in China and accumulates millions of annotated three-component field waveforms. Real-field noise interference and insufficient S-wave energy are ubiquitous across its samples, alongside highly variable source characteristics. Such complicated data layout renders DiTing a proper testbed to quantify model robustness against low-SNR and faint-phase scenarios; we thus employ this dataset exclusively for cross-model comparative analysis.

Unified preprocessing pipelines are enforced across both datasets. Raw seismic recordings are clipped into fixed-length segments under a unified 100 Hz sampling rate, corresponding to 3000 discrete sampling points within a 30 s time window. Instead of rigid one-hot categorical labeling, Gaussian soft labels with 0.1 s standard deviation are adopted. From our training attempts, this labeling scheme smooths label distribution around true onset moments and effectively alleviates training oscillation, and the finalized network outputs three-channel probability sequences matching input length for background, P wave and S wave respectively.

To quantify waveform morphological asymmetry on the same benchmarks used in this study, we evaluate two onset-centered metrics at annotated P and S arrival times. Each three-component segment is normalized with the same zero-mean unit-standard-deviation rule used in training, and an envelope trace is constructed as

$$x(t) = \sqrt{x_Z^2(t) + x_N^2(t) + x_E^2(t)}.$$

Let t_0 denote a candidate arrival time and $L = 50$ samples (0.5 s at 100 Hz). Left and right windows are defined as

$$x_L = x(t_0 - L : t_0), \quad x_R = x(t_0 : t_0 + L),$$

and the mirror-symmetry index is given by

$$\mathcal{S}_{\text{wave}}(t_0) = 1 - \frac{\|x_L - \text{flip}(x_R)\|_2}{\|x_L\|_2 + \|x_R\|_2 + \epsilon},$$

where $\text{flip}(\cdot)$ reverses the sample order and $\mathcal{S}_{\text{wave}} \in [0, 1]$ (higher values indicate stronger mirror symmetry).

To measure asymmetric energy content around the onset, the symmetric window $x(t_0 - L : t_0 + L)$ is decomposed into even and odd components,

$$x_e = \frac{1}{2}(x + \text{flip}(x)), \quad x_o = \frac{1}{2}(x - \text{flip}(x)),$$

and the odd-to-even energy ratio is defined as

$$\mathcal{R}_{eo}(t_0) = \frac{\|x_o\|_2}{\|x_o\|_2 + \|x_e\|_2 + \varepsilon'},$$

with $\mathcal{R}_{eo} \in [0, 1]$ (higher values indicate stronger asymmetry).

We compute both metrics at annotated P and S onsets on the CEED test split and on DiTing samples used for comparative evaluation. Figure 2 summarizes the resulting distributions. On CEED, P arrivals show lower $\mathcal{S}_{\text{wave}}$ and higher $\mathcal{R}_{eo}$ than S arrivals, whereas the same P–S separation remains statistically significant on noisier DiTing recordings (Mann–Whitney U test, $p < 10^{-4}$ in all panels). These results provide a quantitative basis for the morphology-dependent picking asymmetry discussed in Section 2.5 and motivate phase-adaptive feature extraction in SymPhase.

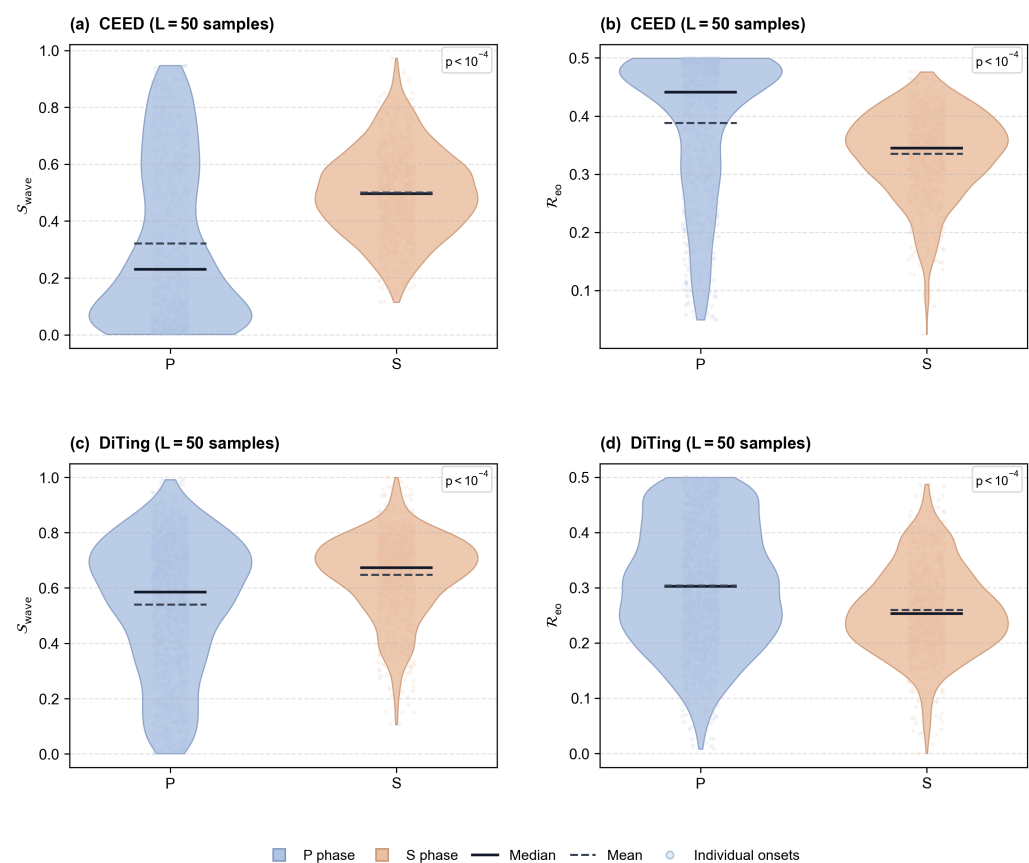


Figure 2. Waveform morphological asymmetry at annotated P and S onsets on the CEED test split and DiTing evaluation samples ($L = 50$ samples, 0.5 s at 100 Hz). Violin plots compare P (light blue) and S (light orange) phase distributions: (a) mirror-symmetry index $\mathcal{S}_{\text{wave}}$ on CEED; (b) odd-to-even energy ratio $\mathcal{R}_{eo}$ on CEED; (c) $\mathcal{S}_{\text{wave}}$ on DiTing; (d) $\mathcal{R}_{eo}$ on DiTing. Solid black and gray horizontal lines mark sample medians and means, respectively; semi-transparent dots show individual onset values. Mann–Whitney U test p -values are annotated in the upper-right corner of each panel.

Onset SNR (dB) is estimated on the same envelope traces using mean-squared amplitudes in a 0.5 s post-onset signal window and a 2.0 s pre-onset noise window (50 and 200 samples at 100 Hz, respectively). Figure 3 shows that P onsets have higher SNR than S onsets on both CEED and DiTing ($p < 10^{-4}$), and that DiTing is shifted toward a lower SNR relative to CEED, indicating stronger field noise while preserving phase-dependent separation.

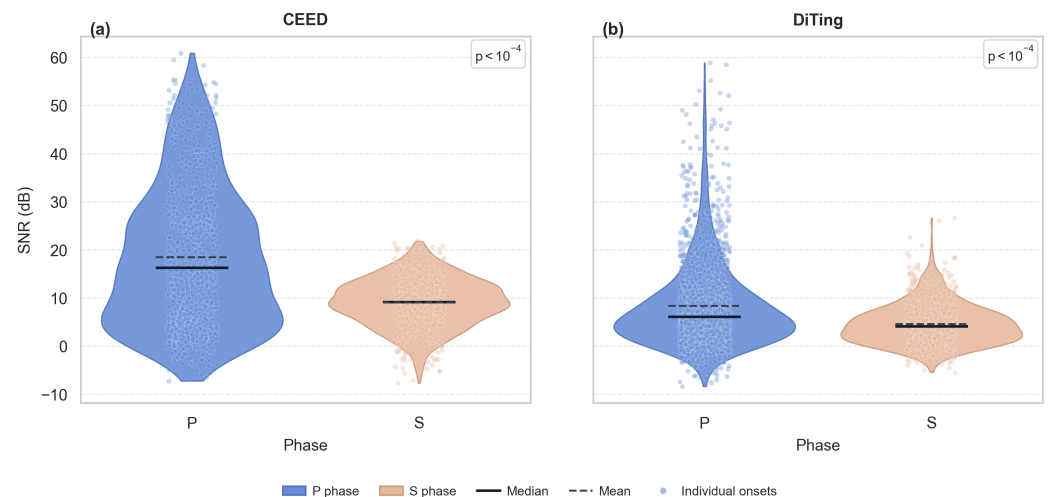


Figure 3. Onset signal-to-noise ratio (SNR, dB) distributions for annotated P and S arrivals on CEED and DiTing. SNR is estimated from a 0.5 s post-onset signal window and a 2.0 s pre-onset noise window on three-component envelope traces (Section 4.1). (a) CEED test split; (b) DiTing evaluation segments. Blue and peach violins denote P and S phases, respectively; solid black and dashed gray horizontal lines indicate medians and means; overlaid dots show individual onset SNR values. Mann–Whitney U test p -values are shown in the upper-right corner of each panel.

4.2. Experimental Setup

4.2.1. Model Configuration and Training

SymPhase is constructed upon PhaseNet-derived 1D U-Net architecture and fully trained using CEED samples. The network takes three-channel seismic input and yields three classification outputs, arranged in a five-layer encoder–decoder layout with the base channel count set to 8, a pooling stride of 4 and dropout coefficient fixed at 0.1. Three disparate kernel sizes (3, 7, 15) constitute multi-scale convolution branches; CBAM and gated skip attention are embedded separately to realize global context refinement and targeted local feature filtering.

SymPhase is trained by minimizing a soft cross-entropy loss over the three-class probability maps (background, P, and S). Let $z_{n,t,c}$ denote the predicted logits and $y_{n,t,c}$ the Gaussian soft label at sample n , time step t , and class c . The training objective is

$$\mathcal{L} = -\frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T \sum_{c=1}^3 y_{n,t,c} \log p_{n,t,c}, \quad p_{n,t,c} = \text{softmax}_c(z_{n,t,\cdot}),$$

where N is the batch size, T is the sequence length, and $p_{n,t,c}$ is the predicted class probability.

AdamW serves as the optimization algorithm. Total training iterations are set to 120 alongside a mini-batch size of 32, with the initial learning rate defined as 5×10^{-3} and weight decay equal to 1×10^{-4} . ReduceLROnPlateau is triggered to shrink the learning rate once validation loss stays unchanged over eight successive epochs, with the lower learning bound constrained at 1×10^{-6} . A constant random seed (2025) is assigned for data division and batch sampling to secure experimental repeatability.

We screen the optimal model checkpoint by referencing composite F1 calculated on the validation dataset:

$$F1_{\text{combined}} = 0.5 \cdot P\text{-F1} + 0.5 \cdot S\text{-F1}.$$

The retained parameter file is loaded for subsequent inference and post-processing operations.

4.2.2. Inference and Evaluation Metrics

After training completion, the network generates probability sequences for three categories during inference, and peak positions extracted from P/S probability curves are defined as predicted arrival instants. Rather than fixing a universal threshold for all samples, we execute independent grid search spanning 0.10 to 0.95 at 0.05 increments on validation set to pick optimal thr_P and thr_S , which are then transplanted onto the test set.

A predicted timestamp is marked correct when its offset against the ground truth stays within 10 sampling points, equivalent to 0.1 s under 100 Hz sampling frequency. We calculate TP, FP, FN and TN sequentially to derive precision, recall and F1 metrics. From practical seismic monitoring experience, excessive false negatives directly lead to missing earthquake records, which explains why F1 is prioritized as the core composite indicator to assess overall picking performance.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

MAE is introduced to quantify temporal prediction discrepancy

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{t}_i - t_i|,$$

where N refers to total count of valid phase arrivals. All experimental schemes run three independent replicates, and final reported metrics adopt averaged values to reduce random experimental fluctuation.

4.3. Ablation Study

We carry out ablation tests on CEED to quantify the individual contribution of each core module embedded inside SymPhase. The plain encoder–decoder network without extra optimization is taken as the baseline (Base). Multi-scale convolution (Multi), gated local selection (Sg) and CBAM global enhancement are added progressively to build four derivative network variants; all hyperparameter configurations remain identical across all training groups. Table 1 lists the mean ablation results on CEED; standard deviations from three repeated trials are supplemented within Appendix A.

Table 1. Ablation experiment results on the California Earthquake Event Dataset (mean values over three runs). The best value in each column is highlighted in **bold**, and the second best is underlined. Prec, Rec, F1, and MAE denote precision, recall, F1 score, and mean absolute error, respectively.

Model	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-MAE (s)	S-MAE (s)
Base	0.9341	0.9636	0.9486	0.8176	0.9340	0.8719	0.1429	0.1522
Base + Multi	0.9627	<u>0.9849</u>	0.9737	<u>0.8638</u>	0.9343	0.8977	0.0640	<u>0.1058</u>
Base + Multi + Sg	0.9666	0.9808	0.9737	0.8624	0.9469	0.9027	<u>0.0603</u>	0.1153
Base + Multi + CBAM	0.9633	0.9847	<u>0.9739</u>	0.8527	0.9674	<u>0.9064</u>	0.0547	0.1098
Base + Multi + CBAM + Sg	<u>0.9635</u>	0.9870	0.9751	0.8655	<u>0.9530</u>	0.9072	0.0686	0.1057

From our tabulated quantitative results, embedding multi-scale convolution brings clear performance advancement over the plain baseline model. When we upgrade the original Base network into the Base + Multi configuration, P-wave F1 rises from 0.9486 to 0.9737, and S-wave F1 reaches 0.8977. Meanwhile, the corresponding MAE values shrink to 0.0640 s for the P phase and 0.1058 s for the S phase. We credit these measurable gains to the multi-branch layout; each branch works on an independent time range to characterize the sharp jump of P onset and the slowly varying energy envelope of S waves individually.

On top of multi-scale modules, adding gated local feature screening yields the Base + Multi + Sg variant. In this setup, P-F1 remains steady at 0.9737, whereas S-F1 further climbs to 0.9027. During our ablation debugging, we observe that the gated component filters useless noisy signals transmitted from skip connections and locks onto locally distributed phase features, which substantially boosts detection precision for low-amplitude, hard-to-identify S waveforms. Figure 4 visualizes these ablation results.

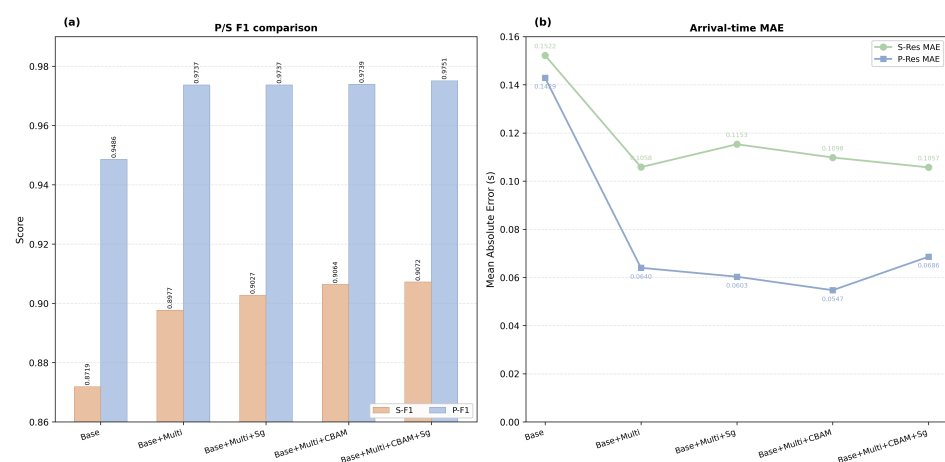


Figure 4. Ablation results on the CEED dataset. (a) P/S-wave F1 scores; (b) arrival-time residual MAE for P- and S-waves on CEED.

Supplementing CBAM global optimization (Base + Multi + CBAM) cuts P-wave MAE down to 0.0547 s and increases S-F1 to 0.9064, benefiting from coordinated channel and temporal feature recalibration for global waveform modeling.

Full combination of three modules yields optimal indicators: P-F1 = 0.9751, S-F1 = 0.9072 with P-MAE = 0.0686 s and S-MAE = 0.1057 s. The collected data validates complementary synergy among three functional components.

4.4. Comparative Experiments

Seven prevailing deep learning pickers are chosen for cross-check to verify SymPhase's comprehensive capability, including TCN, vanilla UNet, Inception, ResASPP, PhaseNet, EQTransformer and EPick. These benchmark models cover distinct technical routes ranging from temporal convolution and classic encoder–decoder to multi-branch atrous structure and convolution–attention hybrid design. All competing algorithms are trained on identical datasets under unified data splitting rules and fixed random seeds; optimal checkpoint selection follows identical validation-set criteria, and the 0.1 s error threshold is universally adopted for arrival correctness judgment.

For DiTing comparative experiments, each benchmark model is trained and evaluated on DiTing using the same fixed H5 split (random seed 2025): the held-out test set contains 800 annotated 30 s three-component segments (8% of a 10,000-sample pool), with the remaining samples used for training and validation. Each segment contributes one sample-level P and one sample-level S decision; a pick is counted as correct if the predicted onset

falls within ± 10 samples (± 0.1 s at 100 Hz) of the manual annotation, consistent with Section 4.2. Descriptive statistics in Figures 2 and 3 use up to 2000 DiTing segments from the same H5 pool.

We evaluate model performance from two independent dimensions. Classification metrics including precision, recall and F1 quantify phase identification capacity, whereas residual mean, standard deviation and MAE reflect temporal positioning precision. Joint assessment of these two sets of metrics delivers comprehensive performance evaluation for all tested algorithms.

4.5. Inference Efficiency and Deployment Considerations

Beyond classification accuracy, computational efficiency is relevant for operational seismic monitoring, where phase pickers must keep pace with continuous waveform streams. We therefore measured pure forward-pass latency on the CEED test split under the same 30 s segments (3000 samples at 100 Hz) and trained checkpoints used in the accuracy comparison (Section 4.2). All timings were obtained on 1000 held-out test waveforms using PyTorch 2.6.0 inference mode. Data loading and checkpoint I/O were excluded. Reported latency includes network forward propagation and softmax normalization over the three output channels; post-processing (P/S probability argmax) required less than 0.004 s for the entire test set on CPU and is therefore omitted from Table 2.

We benchmark inference latency for three representative pickers: SymPhase (our method), PhaseNet (the strongest encoder–decoder baseline on CEED), and TCN (a widely used temporal convolution baseline with a distinct architecture). These models span the proposed design and the two most competitive, structurally distinct routes in our comparison, enabling a direct link between accuracy gains on CEED and computational cost.

GPU experiments used batch size 32 on a single NVIDIA GeForce RTX 4090 (24 GB; driver 580.159.03). CPU experiments used batch size 1 on an Intel Core i9-13900KF workstation (32 logical processors) to approximate single-station, single-stream inference.

Table 2 summarizes all latency results. On GPU, SymPhase achieves 7371 segments/s (0.136 ms per segment). Its throughput is lower than that of PhaseNet (25,811 segments/s) but substantially higher than that of TCN (4548 segments/s). Under the conservative CPU single-stream setting (batch size 1), SymPhase requires 5.32 ms to process one 30 s segment—approximately $3.1\times$ slower than PhaseNet (1.70 ms) yet $2.6\times$ faster than TCN (13.85 ms). Because each segment spans 30 s of continuous observation, even this conservative CPU configuration yields a real-time factor of ~ 5600 for SymPhase, indicating that raw model inference is not the primary performance bottleneck for single-station deployment on a modern desktop CPU.

The main practical obstacles to field deployment are therefore not single-window forward time, but (i) integration with continuous data ingestion and existing monitoring pipelines, (ii) hardware heterogeneity across observation sites, and (iii) operational robustness under 24/7 service. For architectural lightweighting, SymPhase supports depthwise separable convolutions in its multi-scale branches as an optional replacement for standard convolutions. In preliminary tests with all other settings held fixed, this separable variant reduces the parameter count while preserving the same overall topology, but incurs a slight decrease in picking accuracy relative to the standard-convolution configuration. The final SymPhase model reported in this paper therefore adopts standard convolutions to prioritize robustness under low-SNR conditions. Further post-training compression via quantization, structured pruning, and knowledge distillation will be explored in future work to reduce memory footprint and latency on resource-constrained edge nodes without sacrificing the reliability of the full model.

Table 2. Per-segment inference latency and throughput on the CEED test split.

Model	Params (M)	Device	Batch	Latency (ms/seg)	Throughput (seg/s)	RT Factor ^a
TCN	0.984	GPU	32	0.220	4548	1.4×10^5
PhaseNet	0.236	GPU	32	0.039	25,811	7.7×10^5
SymPhase	0.725	GPU	32	0.136	7371	2.2×10^5
TCN	0.984	CPU	1	13.846	72	2166
PhaseNet	0.236	CPU	1	1.695	590	17,700
SymPhase	0.725	CPU	1	5.316	188	5643

^a Real-time factor = 30,000 ms window duration divided by per-segment forward latency (softmax included; post-processing excluded). GPU/CPU settings are specified in Section 4.5.

5. Results

Table 3 aggregates all quantified metrics obtained on the CEED dataset. This benchmark dataset features clean seismic traces and moderate ambient noise, which makes it suitable to inspect the basic capability of competing algorithms. Our core model and its threshold-optimized variant outperform all peer solutions across most evaluation indicators. Raw SymPhase reaches a P-wave F1 of 0.9797 and S-wave F1 of 0.9006, both exceeding PhaseNet's corresponding values. In terms of timing precision, its P-wave MAE drops to 0.0761 s against EQTransformer's 0.1043 s, while S-wave location error is narrowed down from 0.1514 s to 0.1003 s.

Table 3. Model comparison on the CEED dataset (mean values over three runs). SymPhase-G is the best hyperparameter variant of SymPhase obtained via an independent threshold grid search. For each metric, the best value is highlighted in **bold** and the second best is underlined.

Model	Time-Acc	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-Res Mean (s)	P-Res Std (s)	P-Res MAE (s)	S-Res Mean (s)	S-Res Std (s)	S-Res MAE (s)
TCN	0.9936	0.9627	0.9444	0.9534	0.8360	0.9197	0.8758	0.1441	1.6457	0.2839	−0.0076	1.8187	0.2950
UNet	0.9933	0.9514	0.9549	0.9531	0.8161	0.9007	0.8563	<u>0.0102</u>	1.2298	0.1608	−0.0279	1.0209	0.1757
Inception	0.9919	0.9499	0.9413	0.9455	0.7924	0.9488	0.8635	0.1566	2.3241	0.4753	0.0012	1.9981	0.4024
ResASPP	0.9924	0.9485	0.9154	0.9316	0.7892	0.8914	0.8369	0.2077	2.3315	0.5152	0.0235	2.1566	0.4859
PhaseNet	<u>0.9946</u>	<u>0.9658</u>	0.9835	0.9745	<u>0.8443</u>	0.9382	0.8887	0.0190	1.1934	<u>0.1301</u>	−0.0014	1.0543	<u>0.1407</u>
EQTransformer	0.9930	0.9502	0.9702	0.9601	0.8104	0.9366	0.8689	−0.0201	<u>0.8406</u>	<u>0.1043</u>	<u>0.0045</u>	0.9721	0.1514
EPick	0.9934	0.9604	0.9574	0.9589	0.8171	0.8893	0.8516	−0.0027	<u>1.0104</u>	0.1155	−0.0141	1.2705	0.1829
SymPhase	0.9949	0.9707	<u>0.9890</u>	<u>0.9797</u>	0.8516	<u>0.9557</u>	<u>0.9006</u>	<u>−0.0178</u>	0.8063	0.0761	<u>0.0050</u>	0.6180	0.1003
SymPhase-G	0.9949	0.9653	0.9969	0.9808	0.8342	0.9972	0.9084	<u>−0.0178</u>	0.8063	0.0761	<u>0.0050</u>	<u>0.6181</u>	0.1003

After independent threshold fine-tuning on validation data, SymPhase-G gains obvious improvement on recall metrics. Its P recall rises to 0.9969, which is 1.34 percent higher than PhaseNet; meanwhile S recall climbs to 0.9972 with a 4.84 percent increment compared with Inception. Optimized P-F1 and S-F1 hit 0.9808 and 0.9084 respectively, whereas temporal deviation stays identical to original SymPhase. Among all referenced architectures, PhaseNet delivers the second-best overall performance yet still falls behind our proposed design on all key metrics; other mainstream networks such as TCN, Inception and ResASPP lag far behind in both classification accuracy and arrival positioning precision. Figure 5 compares the classification metrics and P-wave MAE on CEED.

Different from CEED, DiTing consists of massive field-collected seismic recordings full of faint S-wave components and heavy ambient interference, which forms a strict testbed to verify model robustness against tough field conditions. All DiTing results in Table 4 are evaluated under the unified ± 0.1 s arrival-time tolerance defined in Section 4.2. On this benchmark, P-wave picking is comparatively less demanding than S-wave picking: P onsets exhibit sharper transients than S onsets (Section 2.5; Figure 2), and annotated P onsets have higher SNR than S onsets even under field noise (Figure 3). As a result, several encoder–decoder pickers (PhaseNet, EQTransformer, EPick, SymPhase, and UNet) attain saturated P-precision, P-recall, and P-F1 of 1.0, with sub-centisecond P-wave MAE well below the 0.1 s tolerance; TCN, Inception, and ResASPP do not reach this ceiling (P-F1 $\approx$ 0.51–0.58). P-wave classification scores therefore lose discriminative power among

strong models on DiTing, and the following comparison focuses on S-phase metrics. Obvious gaps emerge on S-related metrics. Our two variants reach an S-F1 of 0.4724, outperforming PhaseNet's 0.3732 and the 0.28–0.29 range of EQTransformer and EPick; compared with TCN, Inception and ResASPP, the improvement exceeds 0.24 up to 0.31. When it comes to temporal error, SymPhase's S-wave MAE stands at only 0.7386 s, noticeably smaller than EPick's 0.8920 s and PhaseNet's 1.2296 s. UNet almost fails to capture any valid S arrivals, while three conventional convolution-based models suffer from excessive positioning deviation exceeding two seconds. Figure 6 compares the classification metrics and S-wave MAE on DiTing.

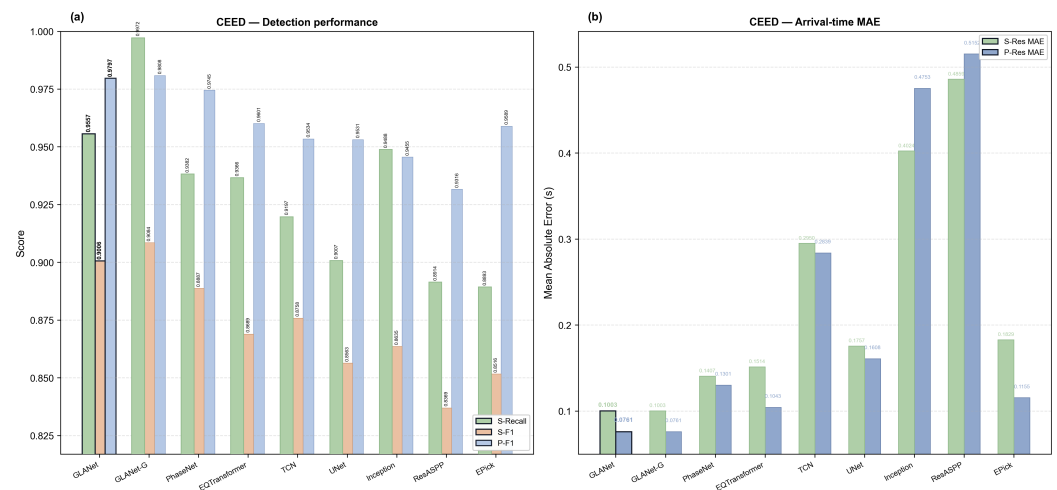


Figure 5. Performance comparison on the CEED dataset. (a) Classification metrics; (b) P-wave arrival-time residual MAE.

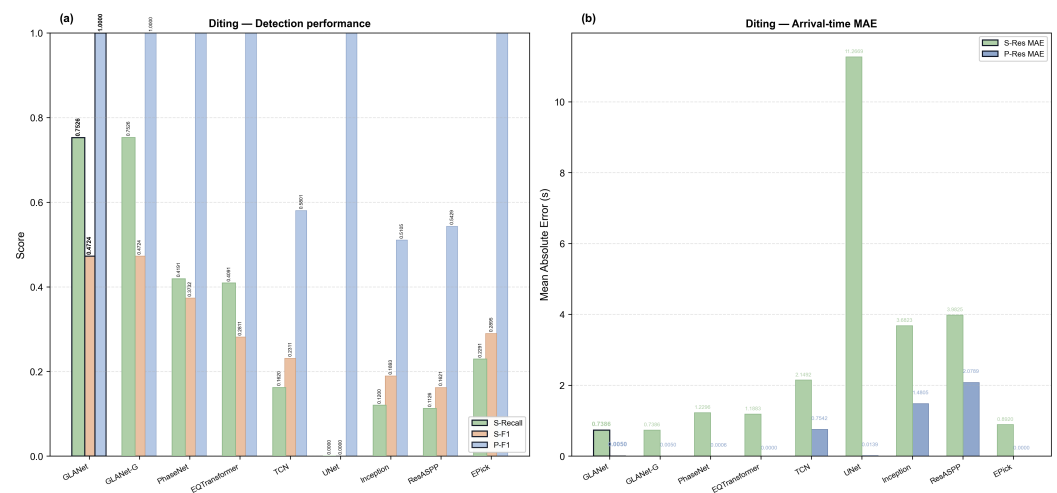
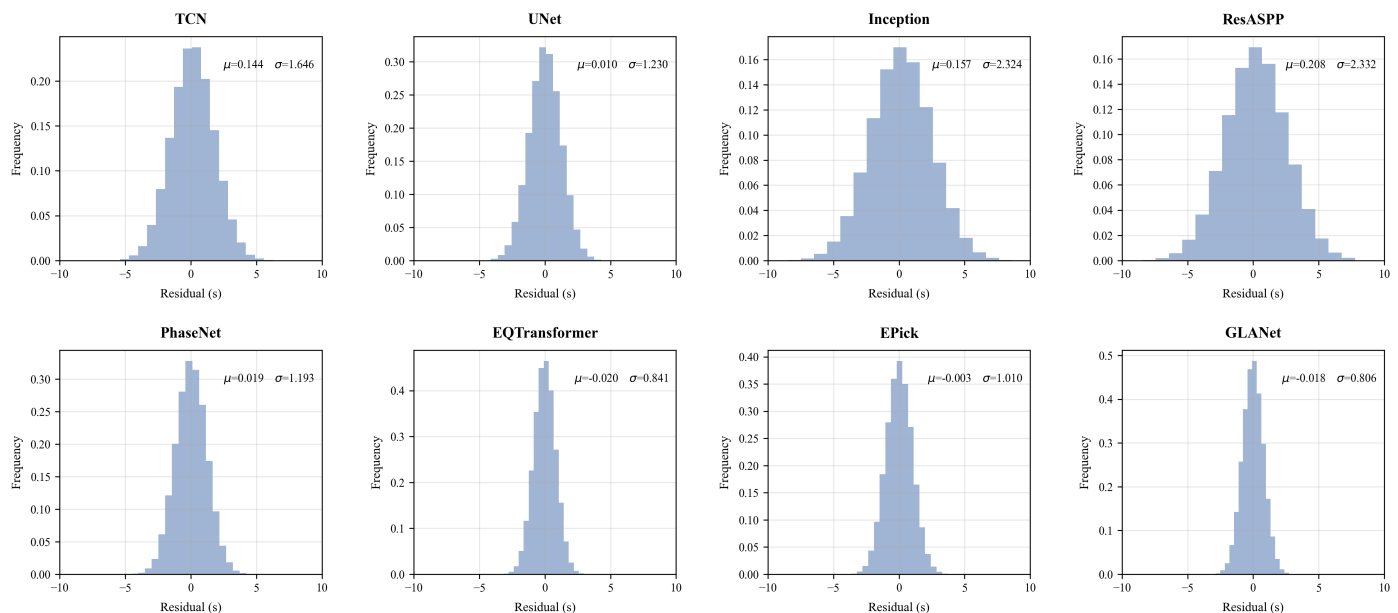


Figure 6. Performance comparison on the DiTing dataset. (a) Classification metrics; (b) S-wave arrival-time residual MAE.

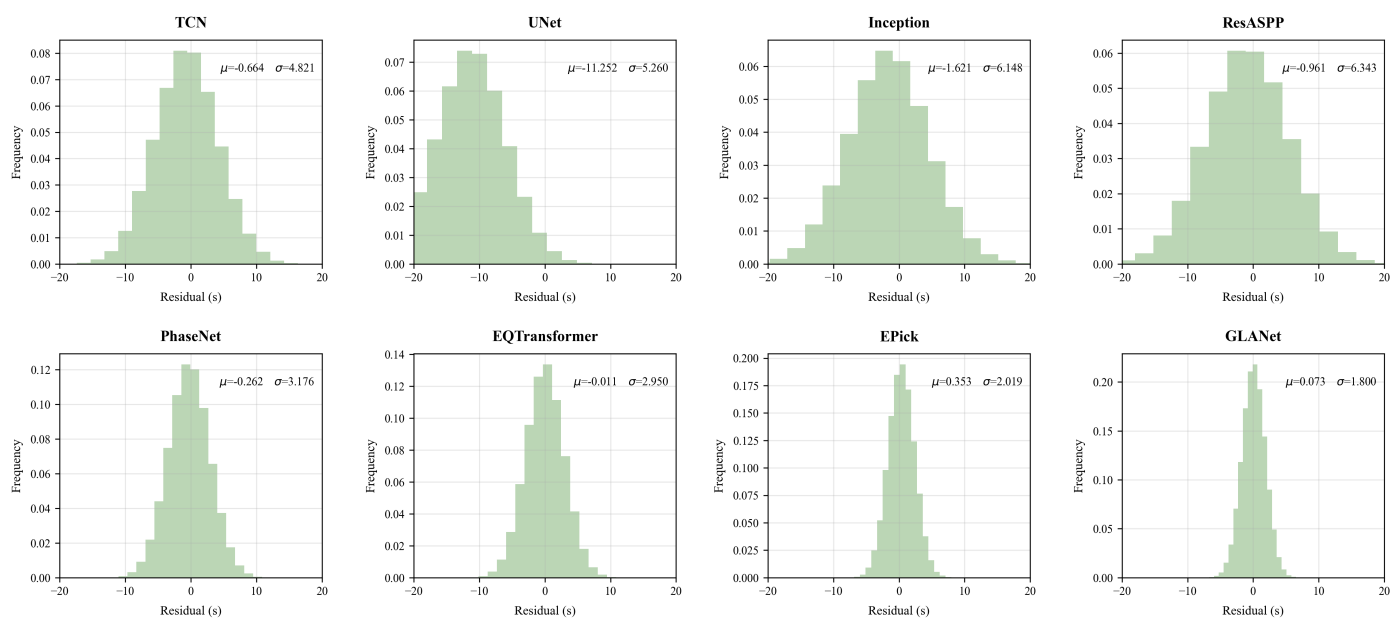
Three core merits can be summarized from the measured data. First, the proposed framework greatly lifts weak S-wave detection accuracy under harsh noisy environments, with F1 improvement ranging from 0.099 to 0.31 against mainstream alternatives. Second, the dual attention design shortens S-wave temporal deviation by 0.15–0.49 s on difficult samples. Third, the whole architecture delivers stable outputs across distinct datasets with small metric fluctuation, demonstrating strong generalization against varied noise and waveform distortion. Figure 7 shows the arrival-time residual distributions.

Table 4. Model comparison on the DiTing dataset (mean values over three runs). SymPhase-G is the best hyperparameter variant of SymPhase obtained via an independent threshold grid search. For each metric, the best value is highlighted in **bold** and the second best is underlined.

Model	Time-Acc	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-Res Mean (s)	P-Res Std (s)	P-Res MAE (s)	S-Res Mean (s)	S-Res Std (s)	S-Res MAE (s)
TCN	0.9890	0.7318	0.4814	0.5801	0.4220	0.1620	0.2311	0.1844	2.1562	0.7542	−0.6637	4.8213	2.1492
UNet	0.9906	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0040	0.0139	−11.2517	5.2603	11.2669
Inception	0.9886	0.7150	0.3802	0.5105	0.3637	0.1200	0.1893	0.3155	3.5045	1.4805	−1.6208	6.1478	3.6823
ResASPP	0.9886	0.6287	0.4815	0.5429	0.3189	0.1126	0.1621	1.0425	4.2863	2.0789	−0.9612	6.3428	3.9825
PhaseNet	0.9943	1.0000	1.0000	1.0000	0.3378	0.4191	0.3732	−0.0005	0.0014	0.0006	−0.2621	3.1756	1.2296
EQTransformer	0.9934	1.0000	1.0000	1.0000	0.2146	0.4091	0.2811	0.0000	0.0000	0.0000	−0.0110	2.9499	1.1883
EPick	0.9946	1.0000	1.0000	1.0000	0.3936	0.2291	0.2895	0.0000	0.0000	0.0000	0.3526	2.0191	0.8920
SymPhase	0.9947	1.0000	1.0000	1.0000	0.3453	0.7526	0.4724	−0.0004	0.0056	0.0050	<u>0.0726</u>	1.7996	0.7386
SymPhase-G	0.9947	1.0000	1.0000	1.0000	0.3453	0.7526	0.4724	−0.0004	0.0056	0.0050	<u>0.0726</u>	1.7996	0.7386



(a) P-wave time residuals



(b) S-wave time residuals

Figure 7. Histograms of arrival-time residuals. (a) P-wave residuals (CEED); (b) S-wave residuals (DiTing). The mean (μ) and standard deviation (σ) are annotated.

Waveform visualization further backs up these numerical findings. Figure 8 compares P/S-wave identification performance on noisy waveforms across four representative models. As displayed in Figure 8, our model generates sharp, well-localized probability peaks even for submerged weak S signals, whereas PhaseNet, TCN and Inception frequently produce blurred curves or entirely miss faint phase responses amid intensive noise. Such intuitive contrasts verify the superiority of our multi-scale plus dual-attention collaborative design in processing complicated field seismic data.

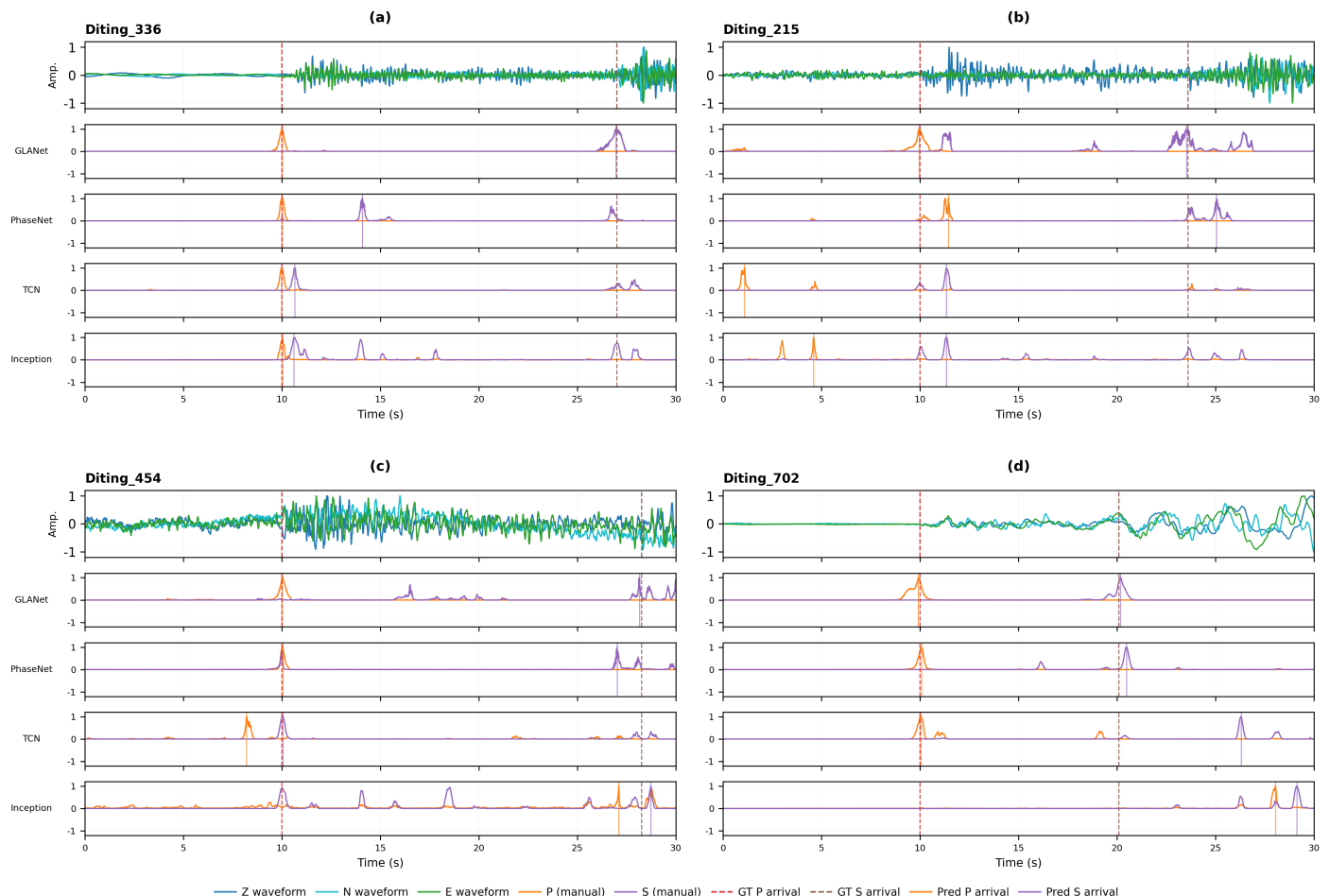


Figure 8. Visualization comparison of P/S picking on four noisy DiTing waveforms across SymPhase, PhaseNet, TCN, and Inception. (a–d) show four representative low-SNR segments. In each panel, the top row plots normalized Z (blue), N (cyan), and E (green) components with dashed red and brown vertical lines marking manual P and S arrivals; lower rows display model-specific P (orange) and S (purple) probability traces, where solid vertical lines indicate predicted onsets and dashed red and brown lines repeat the manual references.

6. Discussion

Quantitative outcomes from ablation and cross-model comparisons explain the source of SymPhase’s competitive performance well. Unlike existing pickers that only focus on either global waveform trends or isolated local details, our network integrates two optimization routes organically to fix inherent structural drawbacks of conventional designs.

Classic single-convolution models lack enough receptive field to cope with nonstationary seismic waveforms, while pure attention networks easily introduce useless noisy features after skip transmission. Benefiting from customized optimization oriented to seismic physical features, SymPhase reserves overall waveform evolution information

and dynamically strengthens abrupt phase-related textures, effectively suppressing oversensitivity to random noise and boosting faint S-wave capture capacity.

The core design philosophy lies in matching symmetric network computation with asymmetric intrinsic waveform properties via triple collaborative modules. The CBAM branch enforces globally unified feature constraint to avoid local noise corruption; gated skip weight brings asymmetric local adjustment to screen useless shallow noise; multi-scale convolution adapts to the divergent temporal span of P and S envelopes. Joint deployment of the above two design modules mitigates conflicts arising from fixed identical network computation and irregular waveform shapes, effectively lowering picking mistakes triggered by inconsistent waveform features.

One existing geophysical study has verified widespread asymmetric traits across various geophysical observation data [39]. The published conclusion coincides with our summed-up rules of seismic waveform variation and adds extra credibility to our symmetry-driven structural design.

Beyond structural improvements inside the network, our separately tuned dual-threshold post-processing also lifts practical picking capability for field measured data. We calibrate P and S thresholds individually using validation datasets; this setup boosts S-wave recall while keeping P-wave precision intact. Since threshold tuning only runs during model inference, all competing models receive identical training conditions and experimental fairness stays guaranteed. This post-processing scheme shortens the gap between network prediction output and the actual field monitoring requirements we encounter in practice.

From the viewpoint of real seismic application, reliable automatic phase picking acts as an essential prerequisite for earthquake relocation and crustal inversion work. Thanks to its robust noise resistance verified in our repeated tests, SymPhase can serve as a usable instrument for automated seismic analysis and help advance fast earthquake early-warning applications.

The inference-efficiency results in Section 4.5 (Table 2) show that SymPhase trades a modest increase in per-segment latency relative to PhaseNet for improved S-phase performance, while remaining substantially faster than TCN under both GPU and CPU settings. For single-station CPU streaming, SymPhase operates orders of magnitude faster than the 30 s acquisition window, suggesting that deployment constraints in operational networks are more likely to arise from system integration and sustained service requirements than from standalone forward-pass time. At the architectural level, depthwise separable convolutions provide a built-in lightweight alternative with fewer parameters, but preliminary tests indicate a slight accuracy penalty relative to the standard-convolution configuration adopted in the final model. Post-training compression techniques remain to be investigated for further deployment on resource-limited edge hardware.

Still, certain limitations remain in the current study. Under extreme ultra-low SNR with extremely weak S energy, occasional phase omission and time deviation can still be observed. In addition, all existing tests are finished on natural earthquake datasets; further field measurement is required to confirm generalization toward severely distorted abnormal waveforms.

Future research directions are defined as follows: post-training compression through quantization, structured pruning, and knowledge distillation for resource-limited edge deployment; advanced cross-dimensional attention for ultra-faint phase extraction; fusion of time-frequency analysis and Transformer structures; plus self-adaptive thresholding and incremental learning to improve on-site adaptability.

7. Conclusions

This study constructs SymPhase, a symmetry-cooperative attention network targeting accurate seismic picking under heavy noise. Equipped with global channel–time attention, gated local refinement and multi-scale adaptive convolution, the built collaborative feature learning architecture outperforms PhaseNet, EQTransformer and other mainstream alternatives on both CEED and tough DiTing datasets, with prominent gain on low-SNR weak S-wave identification.

Different from traditional equalized feature extraction schemes, our framework bridges the gap between symmetric network computation and inherently asymmetric seismic waveforms via global constraint plus local adaptive weighting, reducing missing picks and temporal deviation efficiently when combined with dual-threshold post-processing.

In summary, SymPhase delivers a feasible solution for high-noise automatic phase picking. The proposed global–local collaborative learning scheme provides a new route for seismic time-series analysis and bears application potential in automatic seismic monitoring, source inversion and geohazard prevention.

Author Contributions: Conceptualization, Y.W. and J.W.; Methodology, Y.W.; Software, Y.W. and Q.L.; Validation, Y.W. and C.Z.; Formal Analysis, C.Z.; Investigation, Y.W., Q.L. and Y.B.; Data Curation, Q.L. and Y.B.; Visualization, Q.L. and Y.B.; Writing—Original Draft Preparation, Y.W.; Writing—Review and Editing, Y.H., C.Z., X.D. and J.W.; Supervision, J.W. and Y.H.; Project Administration, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The CEED (<https://huggingface.co/datasets/AI4EPS/CEED>, accessed on 3 June 2026) and DiTing (<https://data.earthquake.cn/datashare/report.shtml?PAGEID=datasourcelist&dt=ff8080827e3317f4017e331e7ce00002>, accessed on 3 June 2026) seismic datasets utilized in this study are publicly available. Researchers can access these data via their official repositories and the cited literature for academic purposes [37,38]. The code used in this study is publicly available at GitHub: <https://github.com/nanxun00/SymPhase> (accessed on 3 June 2026).

Acknowledgments: Use of AI Tools Statement: During the preparation of this manuscript, the author used AI tools solely for English language editing and grammar correction. The author reviewed and edited all AI-generated outputs and takes full responsibility for the accuracy, integrity, and originality of the content.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SymPhase	Symmetry-Aware Collaborative Attention Network
SNR	Signal-to-Noise Ratio
CEED	California Earthquake Event Dataset
MAE	Mean Absolute Error
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative

Appendix A. Standard Deviation Tables

Table A1. Standard deviations of the ablation metrics on the CEED dataset (over three runs).

Model	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-MAE (s)	S-MAE (s)
Base	0.0022	0.0098	0.0058	0.0095	0.0158	0.0119	0.0258	0.0096
Base + Multi	0.0023	0.0006	0.0012	0.0048	0.0094	0.0044	0.0170	0.0125
Base + Multi + Sg	0.0012	0.0055	0.0028	0.0102	0.0063	0.0030	0.0094	0.0082
Base + Multi + CBAM	0.0021	0.0030	0.0021	0.0084	0.0036	0.0038	0.0045	0.0161
Base + Multi + CBAM + Sg	0.0026	0.0052	0.0023	0.0061	0.0101	0.0063	0.0284	0.0052

Table A2. Standard deviations of model comparison metrics on the CEED dataset (over three runs).

Model	Time-Acc	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-Res Mean	P-Res Std	P-Res MAE	S-Res Mean	S-Res Std	S-Res MAE
TCN	0.0001	0.0039	0.0052	0.0039	0.0018	0.0141	0.0054	0.0289	0.1704	0.0353	0.0191	0.0932	0.0166
UNet	0.0000	0.0001	0.0084	0.0041	0.0022	0.0162	0.0085	0.0050	0.0370	0.0125	0.0413	0.2590	0.0257
Inception	0.0006	0.0031	0.0035	0.0006	0.0109	0.0013	0.0061	0.0870	0.2438	0.0560	0.0114	0.0552	0.0167
ResASPP	0.0001	0.0069	0.0082	0.0033	0.0098	0.0247	0.0081	0.0086	0.2328	0.0685	0.0357	0.0470	0.0229
PhaseNet	0.0001	0.0040	0.0026	0.0010	0.0061	0.0075	0.0080	0.0064	0.3176	0.0130	0.0291	0.2587	0.0305
EQTransformer	0.0003	0.0045	0.0075	0.0060	0.0054	0.0143	0.0036	0.0304	0.4219	0.0399	0.0079	0.2231	0.0156
EPick	0.0001	0.0011	0.0062	0.0015	0.0129	0.0097	0.0026	0.0288	0.1625	0.0085	0.0319	0.1607	0.0136
SymPhase	0.0001	0.0026	0.0012	0.0011	0.0104	0.0117	0.0023	0.0276	0.3349	0.0151	0.0228	0.4015	0.0233
SymPhase-G	0.0001	0.0050	0.0042	0.0006	0.0039	0.0024	0.0034	0.0279	0.3304	0.0144	0.0228	0.4277	0.0238

Table A3. Standard deviations of model comparison metrics on the DiTing dataset (over three runs).

Model	Time-Acc	P-Prec	P-Rec	P-F1	S-Prec	S-Rec	S-F1	P-Res Mean	P-Res Std	P-Res MAE	S-Res Mean	S-Res Std	S-Res MAE
TCN	0.0002	0.0329	0.0184	0.0154	0.0109	0.0541	0.0555	0.0886	0.3423	0.0905	0.1039	0.2947	0.1002
UNet	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0133	0.0038	0.0101	0.2311	0.0779	0.1859
Inception	0.0003	0.0404	0.0265	0.0280	0.0835	0.0317	0.0194	0.3747	0.1587	0.1608	0.7427	0.4013	0.4254
ResASPP	0.0001	0.0065	0.0081	0.0032	0.0097	0.0245	0.0080	0.0085	0.2327	0.0684	0.0356	0.0469	0.0228
PhaseNet	0.0004	0.0000	0.0000	0.0000	0.0258	0.0197	0.0150	0.0008	0.0019	0.0008	0.3522	0.6797	0.2180
EQTransformer	0.0004	0.0000	0.0000	0.0000	0.0143	0.0634	0.0271	0.0000	0.0000	0.0000	0.4299	1.1899	0.2011
EPick	0.0004	0.0000	0.0000	0.0000	0.0713	0.0400	0.0517	0.0000	0.0000	0.0000	0.1306	0.2630	0.0746
SymPhase	0.0002	0.0000	0.0000	0.0000	0.0219	0.0941	0.0331	0.0007	0.0097	0.0087	0.1666	0.0996	0.0106
SymPhase-G	0.0002	0.0000	0.0000	0.0000	0.0221	0.0932	0.0330	0.0007	0.0097	0.0087	0.1621	0.0999	0.0103

References

- Allen, R.V. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* **1978**, *68*, 1521–1532. [\[CrossRef\]](#)
- Akazawa, T. A Technique for Automatic Detection of Onset Time of P- and S-Phases in Strong Motion Records. In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004; Paper No. 786.
- Baer, M.; Kradolfer, U. An automatic phase picker for local and teleseismic events. *Bull. Seismol. Soc. Am.* **1987**, *77*, 1437–1445. [\[CrossRef\]](#)
- Sleeman, R.; Van Eck, T. Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings. *Phys. Earth Planet. Inter.* **1999**, *113*, 265–275. [\[CrossRef\]](#)
- Zhu, W.; Beroza, G.C. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* **2019**, *216*, 261–273. [\[CrossRef\]](#)
- Mousavi, S.M.; Ellsworth, W.L.; Zhu, W.; Chuang, L.Y.; Beroza, G.C. Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **2020**, *11*, 3952. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, W.; Chakraborty, M.; Fenner, D.; Faber, J.; Zhou, K.; Rumpker, G.; Stöcker, H.; Srivastava, N. EPick: Attention-based multi-scale UNet for earthquake detection and seismic phase picking. *Front. Earth Sci.* **2022**, *10*, 953007. [\[CrossRef\]](#)
- Zhao, S.; Wang, J.; Huang, P.; Zhao, F.; Zhang, F.; Zhu, Y. A multi-scale feature fusion network based on semi-channel attention for seismic phase picking. *Eng. Appl. Artif. Intell.* **2025**, *141*, 109739. [\[CrossRef\]](#)
- Song, D.; Dai, F.; Liu, Y.; Wei, M.; Tan, H. A multi-scale CNN-Transformer hybrid network for microseismic signal arrival picking: Model analysis and engineering application. *Eng. Geol.* **2025**, *353*, 108109. [\[CrossRef\]](#)
- Perol, T.; Gharbi, M.; Denolle, M. Convolutional neural network for earthquake detection and location. *Sci. Adv.* **2018**, *4*, e1700578. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, Y. Automatic microseismic event picking via unsupervised machine learning. *Geophys. J. Int.* **2020**, *222*, 1750–1764. [\[CrossRef\]](#)

12. Zhai, S.B.; Cheng, J.M.; Li, G.W.; Wang, Z.Y.; Xing, Y.M.; Hou, M.S. Robust auto-landing guidance and control for fixed-wing unmanned aerial vehicles with strong wind disturbances. *Aerosp. Sci. Technol.* **2026**, *178*, 112413. [\[CrossRef\]](#)
13. Zhai, S.B.; Li, G.W.; Wu, G.; Nan, K.Z.; Zhang, H.Z.; Hou, M.S. Cooperative 4D trajectory planning for multi aerial vehicles combining receding horizon optimization and differential evolution algorithm in dynamic battlefield environment. *Appl. Soft Comput.* **2026**, *197*, 115169. [\[CrossRef\]](#)
14. Wu, X.; Dong, J.; Bao, W.; Zou, B.; Wang, L.; Wang, H. Augmented intelligence of things for emergency vehicle secure trajectory prediction and task offloading. *IEEE Internet Things J.* **2024**, *11*, 36030–36043. [\[CrossRef\]](#)
15. Chen, D.; Xu, C.; Yu, J.; Wang, Q.; Fang, H.; Yin, S.; Lookman, T.; Chen, R. Interpretable Machine Learning Framework for Nb–Si Based Alloy Design with Enhanced Fracture Toughness. *Adv. Sci.* **2026**, e75815. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Chen, J.; Shao, Z.; Zhu, H.; Chen, Y.; Li, Y.; Zeng, Z.; Yang, Y.; Wu, J.; Hu, B. Sustainable interior design: A new approach to intelligent design and automated manufacturing based on Grasshopper. *Comput. Ind. Eng.* **2023**, *183*, 109509. [\[CrossRef\]](#)
17. Chen, J.; Yin, H.; Zhang, K.; Ren, Y.; Zeng, H. Integration of neural networks in brain–computer interface applications: Research frontiers and trend analysis based on Python. *Eng. Appl. Artif. Intell.* **2025**, *151*, 110654. [\[CrossRef\]](#)
18. Wu, X.; Zou, B.; Lu, C.; Wang, L.; Zhang, Y.; Wang, H. Dynamic Security Computing Framework With Zero Trust Based on Privacy Domain Prevention and Control Theory. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 2266–2278. [\[CrossRef\]](#)
19. Wu, X.; Wang, H.; Wei, D.; Shi, M. ANFIS with natural language processing and gray relational analysis based cloud computing framework for real time energy efficient resource allocation. *Comput. Commun.* **2020**, *150*, 122–130. [\[CrossRef\]](#)
20. Chen, L. Beyond external constraints: The missing dimension of AI governance. *SSRN Electron. J.* **2026**. [\[CrossRef\]](#)
21. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 3–19. [\[CrossRef\]](#)
22. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**. [\[CrossRef\]](#)
23. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 833–851. [\[CrossRef\]](#)
24. Pardo, E.; Garfias, C.; Malpica, N. Seismic phase picking using convolutional networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7086–7092. [\[CrossRef\]](#)
25. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**. [\[CrossRef\]](#)
26. Han, Z.; Li, Y.; Guo, K.; Li, G.; Zheng, W.; Liu, H. A Seismic Phase Recognition Algorithm Based on Time Convolution Networks. *Appl. Sci.* **2022**, *12*, 9547. [\[CrossRef\]](#)
27. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [\[CrossRef\]](#)
28. Li, G.; Yuan, M.; Li, Y.; Huang, J.; Zhang, L.; Zhang, H.; Han, Z. Seismic P-wave first-arrival picking model based on EQK-IncResNet. *Concurr. Comput. Pract. Exp.* **2023**, *35*, e7392. [\[CrossRef\]](#)
29. Liao, W.Y.; Lee, E.J.; Mu, D.; Chen, P.; Rau, R.J. ARRU phase picker: Attention recurrent-residual U-Net for picking seismic P- and S-phase arrivals. *Seismol. Res. Lett.* **2021**, *92*, 2410–2428. [\[CrossRef\]](#)
30. Wang, S.; Liu, F.; Yin, X.X.; Chen, K.; Cai, R. Employing convolution-enhanced attention mechanisms for earthquake detection and phase picking models. *Front. Earth Sci.* **2023**, *11*, 1283857. [\[CrossRef\]](#)
31. Yu, Z.; Wang, W.; Chen, Y. Benchmark on the accuracy and efficiency of several neural network based phase pickers using datasets from China Seismic Network. *Earthq. Sci.* **2023**, *36*, 113–131. [\[CrossRef\]](#)
32. Mousavi, S.M.; Beroza, G.C. Machine Learning in Earthquake Seismology. *Annu. Rev. Earth Planet. Sci.* **2023**, *51*, 105–129. [\[CrossRef\]](#)
33. Yu, Z.; Wang, W. LPPN: A lightweight network for fast phase picking. *Seismol. Res. Lett.* **2022**, *93*, 2834–2846. [\[CrossRef\]](#)
34. Li, B.; Fan, L.; Jiang, C.; Liao, S.; Fang, L. CSESnet: A deep learning P-wave detection model based on UNet++ designed for China Seismic Experimental Site. *Front. Earth Sci.* **2023**, *10*, 1032839. [\[CrossRef\]](#)
35. Zhang, Y.; Liu, X.Q.; Song, L.W.; Dong, H.L. Microseismic event detection based on multiscale detection convolutional neural network. *J. Seism. Explor.* **2023**, *32*, 455–477.
36. Li, Y.; Wang, Z.; Wang, J.; Sui, Q.; Li, S.; Wang, H.; Cao, Z. First Arrival Picking on Microseismic Signals Based on K-Means with a ReliefF Algorithm. *Symmetry* **2021**, *13*, 790. [\[CrossRef\]](#)
37. Zhu, W.; Wang, H.; Rong, B.; Yu, E.; Zuzlewski, S.; Tepp, G.; Taira, T.; Marty, J.; Husker, A.; Allen, R.M. California Earthquake Dataset for Machine Learning and Cloud Computing. *arXiv* **2025**. [\[CrossRef\]](#)

38. Zhao, M.; Xiao, Z.; Chen, S.; Fang, L. DiTing: A large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthq. Sci.* **2023**, *36*, 84–94. [[CrossRef](#)]
39. Sutyurin, G.G. The Importance of Being Asymmetric for Geophysical Vortices. *Symmetry* **2023**, *15*, 2204. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.