



TranSeis: A high precision multitask seismic waveform detector

Yuxin Zhou ^{a,b,c}, Huai Zhang ^{a,b,*}, Shi Chen ^d, Zheng Yuan ^c, Chuanqi Tan ^c, Fei Huang ^{c,**}, Yicun Guo ^{a,b}, Yaolin Shi ^{a,b}

^a State Key Laboratory of Earth System Numerical Modeling and Application, Chinese Academy of Sciences, Beijing, 100049, China

^b College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

^c Alibaba Group, Hangzhou, 311121, China

^d Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China

ARTICLE INFO

Keywords:

Seismic waveform data

Multitask learning

Multi-GPU parallel processing

Artificial intelligence

ABSTRACT

This work introduces a highly efficient multitask parallel Artificial Intelligence model designed for weak seismic signal detection and phase picking, leveraging the capabilities of the conventional AI-powered Transformer architecture. By integrating a multi-part data extraction strategy, a multi-GPU parallel processing framework, and a multi-layer network schedule, we significantly enhance the accuracy of detecting P- and S-phases while optimizing the model's efficiency. The accuracy attained for the P and S phases was 92% and 76% when employing only a segment of the dataset. When we incorporated the entire dataset, the precision improved to 97% for P phases and 87% for S phases. Notably, our model demonstrates higher accuracy compared to existing deep-learning and traditional detection algorithms. When applied to extensive seismic phase observation data collected from 2020 to 2023 in mainland China, our model consistently demonstrated high accuracy, confirming its generalizability across various spatiotemporal contexts. It also exhibited exceptional sensitivity to subtle changes in waveform data, highlighting its promising potential for detecting smaller seismic events with even greater resolution in future applications.

1. Introduction

The data-driven approach to exploring the characteristics of seismic activity provides a promising way to predict significant earthquake events, however it also poses challenges for the geoscience community. Consequently, numerical earthquake prediction has emerged as a frontier in the field, with its core being the quantitative probabilistic estimation of spatiotemporal patterns and magnitudes through large-scale computational simulations and big data-driven artificial intelligence. A key process in numerical earthquake prediction is the extraction of features and the identification of anomalies directly from seismic waveforms.

Essentially, the detection of earthquake precursors and earthquake prediction based on standard geophysical fields involves three key components: first, the detection of weak signals. Second, the identification of anomalous signals; and the foundation for distinguishing anomalous signals is the establishment of a standard geophysical field.

Finally, earthquake precursor information is hidden in anomalous signals that are separated from standard signals.

This process raises two significant scientific issues: (a) How to achieve reliable identification and enhancement of weak signals. (b) How to select an appropriate deep learning architecture to extract dependable precursor information.

The essence of weak signal detection lies in the extraction of features from seismic waveforms, which might be challenging to detect using traditional methods, especially given the exponential rise of observational data acquired from multiple seismic stations. Because traditional seismic signal detection methods typically rely on the individual expertise and subjective judgment of professionals. For example, the STA/LTA method, which is based on the ratios of short-term average and long-term average (Allen, 1978; Baer and Kradolfer, 1987; Withers et al., 1998). As a result, high-precision phase detection and accurate phase picking of seismic waveforms that incorporate AI-powered methods have merged as priority tasks in advanced seismological investigations

* Corresponding author. State Key Laboratory of Earth System Numerical Modeling and Application, Chinese Academy of Sciences, Beijing, 100049, China.

** Corresponding author.

E-mail addresses: zhouyuxin21@mails.ucas.ac.cn (Y. Zhou), hzhang@ucas.ac.cn (H. Zhang), chenshi@cea-igp.ac.cn (S. Chen), mike29@126.com (Z. Yuan), ysjtcq@gmail.com (C. Tan), f.huang@alibaba-inc.com (F. Huang), guoyicun@ucas.ac.cn (Y. Guo), shiyi@ucas.ac.cn (Y. Shi).

to date, serving as one of the primary motivations for our research.

Deep learning (LeCun et al., 2015) models have demonstrated powerful feature extraction capabilities in image and sequence data processing. Ross et al. (2016) have applied Convolutional Neural Networks (CNNs) to the field of earthquakes. Perol et al. (2018) utilized the CNN network to process tasks such as seismic signal and noise classification, as well as seismic phase classification. Experiments have shown that the CNN method is feasible, while in terms of seismic phase time picking, the accuracy was still insufficient compared to traditional methods and manual processing results (Perol et al., 2018; Ross et al., 2016). Ronneberger et al. (2015) proposed a U-net based on fully convolutional networks (FCNN) (Long et al., 2015) and achieved results similar to human accuracy in medical image segmentation (Long et al., 2015; Ronneberger et al., 2015). Zhao et al. (Ming et al., 2019) applied U-Net to the earthquake phase arrival dataset recorded by the Wenchuan earthquake aftershocks and the Capital Area Seismic Network. Zhu et al. (Zhu and Beroza, 2018) proposed PhaseNet, a deep neural network algorithm for seismic phase picking. The network learns features directly from labeled data, offering significant improvements over traditional STA/LTA methods. In 2020, Mousavi et al. (2020) used the transformer network with the attention mechanism to train and test approximately 1.2 million waveform data globally, known as EQTransformer. However, the performance of EQTransformer has decreased in the Yang Bi and Madao earthquake aftershock sequences in China (Jiang et al., 2021). Similar results were observed in the 2019 Ms 5.4 earthquake sequence in Wenyuan, Sichuan, China. EQT has a limited recall rate, and the number of detected earthquakes is lower than PhaseNet (Lin Xuekai, 2022). On the one hand, China's complex and unique topographical features, distinct from those of other countries, and the original training data for the existing models do not include Chinese data, which limits the ability of pre-trained models to effectively learn from Chinese seismic data. On the other hand, while direct transfer learning can partially ensure the practical applicability of pre-existing models in new regions, differences in data types and data collection methods necessitate adjustments of models. Such adjustments may unavoidably compromise the original model's learning capacity. Therefore, one of the key motivations for this work is to develop a highly interpretable picker algorithm that ensures high performance on seismic data from mainland China and can maintain generalizability across different regions within the country.

In this study, taking advantage of the theoretical strengths of the Transformer architecture in processing time-series data, we designed a high-precision neural network, TranSeis, specifically tailored for seismic data from mainland China. The model employs a multi-task structure, incorporating a parallel processing mode for multiple parts of the data, and it was trained and tested on a large-scale dataset from earthquake events across mainland China. Additionally, the model was evaluated on different datasets and continuous waveforms from regions across the mainland, demonstrating its robustness and adaptability.

2. Data and methods

In this paper, we use a large-scale nationwide dataset called DiTing to train the model, released by the National Seismic Data Centre (NSDC) in 2023 and collected by Zhao et al. The events are sourced from the earthquake observation reports issued by the China Earthquake - Networks Center from 2013 to 2020. Following meticulous data cleansing and desensitization procedures, the dataset comprises 787,010 seismic events meticulously recorded by over 1300 fixed seismic stations distributed throughout China with epicenter distances of 330 km, corresponding to 2,734,748 three-component seismic waveforms, each meticulously annotated with P and S phase labels. All waveforms in the dataset are uniformly sampled at a frequency of 100 Hz. Each waveform begins with a random duration of seconds before the earthquake event and is cut into a standardized length of 200 s.

The DiTing dataset boasts a significantly larger sample size (see Fig. 1), and its distribution range of magnitude and signal-to-noise ratio is more uniform compared to previous large-scale datasets such as SCSNDC (Ross et al., 2018), STEAD (Mousavi et al., 2019), and INSTANCE (Michelini et al., 2021). More importantly, DiTing is a labeled data set based on the waveforms of China's seismic network, which records the distribution of statistical features in different areas of mainland China, reflecting the complex crustal structure and the seismic activity patterns. Using DiTing for training can help develop a seismic data processing model tailored for the Chinese mainland and, to some extent, ensure the generalizability of the model in different regions of China.

We bring forward a supervised learning paradigm, where the input seismic waveforms are labeled with the corresponding signals. In this paper, we set the sample points starting from the P-wave arrival time to the S-wave arrival time + c × (S-P) as 1 and the rest as 0, with c as an optional parameter, or we directly set the P-wave arrival time until the end as 1 and the rest of the positions as 0, and take the vector characterized as (20000, 1) as the target label for seismic signal detection. We use one-hot coding for P and S labeling.

Our proposed TranSeis (see Fig. 2) combines CNN, RNN, LSTM, Transformer, and other neural network modules. It inherits the multi-task structure of EQTransformer and uses a three-branch mode in the decoder module to complete signal detection and P-S seismic phase picking simultaneously. . The encoder-decoder architecture is employed, with the encoder primarily consisting of convolutional and max-pooling layers. Convolution layers, which are widely used in image recognition (He et al., 2016; Ioffe and Szegedy, 2015; Krizhevsky et al., 2012; Simonyan and Zisserman, 2014), extract image features from input images. Similar to image recognition, waveform data also encapsulates features across three channels, manifesting through the combined or independent representation of each sampling point.

Our model starts with an encoder and two layers of RNN modules. The convolutional layer in the residual block is preceded by a batch normalization layer, an activation layer, and a dropout layer (Srivastava et al., 2014). The integration of normalization and dropout mechanisms aims to accelerate training and reduce the risks of overfitting. Subsequently, the network progresses through a BiLSTM module (Huang et al., 2015; Zhou et al., 2016). They consist of a BiLSTM layer followed by a convolutional layer, effectively deepening the network without increasing the number of learnable parameters. Notably, the LSTM layers, before self-attention layers, are indispensable for incorporating positional information (Chan et al., 2016; Sperber et al., 2018).

The positional information vector is then amalgamated with the sequence's representation vector, serving as the input to the Transformer module \mathbf{x} . The output of the Transformer template is calculated according to

$$\alpha_t = \text{LayerNorm}(\text{SeqSelfAttention}(\mathbf{x}) + \mathbf{x}), \quad (1)$$

$$\mathbf{F}_t = \max(\mathbf{0}, \alpha_t \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2, \quad (2)$$

$$\alpha_t' = \text{LayerNorm}(\mathbf{F}_t + \alpha_t) \quad (3)$$

where , α_t represents the output of the first normalized layer, where the input of neurons in each layer is converted into the uniform mean and variance to accelerate convergence. \mathbf{F}_t denotes the output of the Feed Forward layer, consisting of two fully connected layers with 128 hidden units each. The first layer is activated using ReLu, while the second layer is not activated. α_t' represents the ultimate output after residual connection and layer normalization, serving as the final output of the Transformer module.

In the task of seismic signal detection, the output of the transformer module will be directly connected to decoders. Decoders and encoders exhibit symmetry, where the fundamental network layer consists of an up-sampling layer (Luo et al., 2023) followed by a convolutional layer.

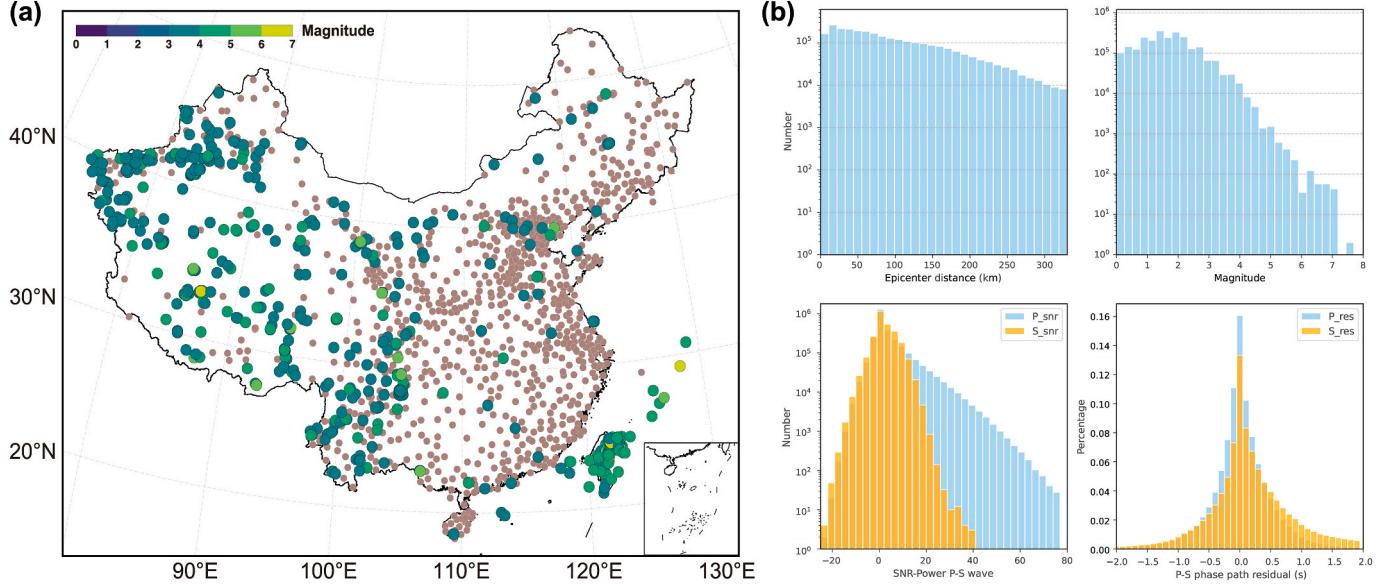


Fig. 1. (a) The locations of more than 1300 seismic stations (brown points) and selected seismic events in training datasets. The distribution and the magnitude of seismic events are shown using the color bar, and the magnitude of seismic events ranges from 0 to 7.7; (b) the distribution of the epicentral distance, the distribution of magnitude, and the distribution of the P-wave and S-wave S/N ratios. The P-wave S/N ratios are mainly distributed in the range of -0.05 to 5.31 dB, and the S-wave S/N ratios are mainly distributed in the range of -0.05 to 4.73 dB.

Upsampling amplifies the output size of the transformer block, and a cropping layer is added to every set of three fundamental network layers. After seven iterations of these decoder operations, the output dimension is determined by a convolutional layer, and the sigmoid function is used to activate the probability of each sample point in the output.

In contrast to seismic signal detection, where the emphasis lies on capturing the onset of signals, the picking of P and S phases places greater emphasis on partially truncated waveform segments rather than full waveform data. We propose leveraging the attention mechanism to spotlight waveform segments during the attenuation phase of seismic signals. It involves integrating an LSTM layer preceding the decoder module and then applying the local attention mechanism to selectively focus on specific segments of the sequence, thereby enhancing the model's capability to discern individual seismic phases within the waveform.

3. Model training strategy

All datasets include header files in CSV format and waveform data files in HDF5 format. Initially, the entire dataset is evenly divided into 28-part files, each containing 100,000 three-component waveforms along with their corresponding metadata. Subsequently, the data within each partition is randomly divided into training, validation, and test subsets, comprising 60%, 20%, and 20% of the total data, respectively. These subsets are sequentially aggregated to form the complete training set (1,640,850 three-component waveforms in total), validation data (546,949 in total), and test data (546,949 in total). During each training iteration, batch size/28 pieces of data is extracted from the training data with a step length of 60,000. In essence, this involves the simultaneous random extraction of data from all 28 partitioned files to create a batch of training data, constituting an initial form of parallel data reading (see Fig. 3). The challenge of processing large-scale data has become a significant obstacle encountered in deep learning. With the increasing volume of data, the demand for memory resources rises, and the speed of data reading becomes a critical concern. To address these challenges, we employed a multi-card simultaneous reading approach for a portion of the dataset. This approach was aimed at alleviating memory overflow issues during experimentation and reducing data reading time

(Rajbhandari et al., 2020; Shoeybi et al., 2019).

The waveform is labeled in Section 2, and the network is trained with labeled samples. After the deep neural network module, the output layer uses the sigmoid function to predict the probability of each sample point being the arrival time of seismic waves. The maximum probability is then taken as the P-wave arrival time and S-wave arrival time predicted by the model. Then, the binary cross-entropy loss between the predicted results and the ground truth labels is calculated, and the Adam optimization algorithm (Kingma and Ba, 2014) is employed to optimize the computational process.

Model performance is evaluated using metrics such as precision, recall, and F1 score, which provide a comprehensive assessment of the model's effectiveness in seismic signal detection and phase pickup tasks (see Fig. 4). The predicted result is considered a true positive (TP) if the absolute distance between the predicted sample points and the ground truth sample points is less than 0.5 s (i.e., not more than 500 sample points).

- Precision

$$\text{precision} = \frac{TP}{TP + FP} \quad (4)$$

- Recall

$$\text{recall} = \frac{TP}{TP + FN} \quad (5)$$

- F1 score

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

If preferences for recall and precision differ, the different emphasis on recall or precision can be expressed by adjusting the parameter β , as illustrated by the following formula:

$$F_\beta = \frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}} \quad (7)$$

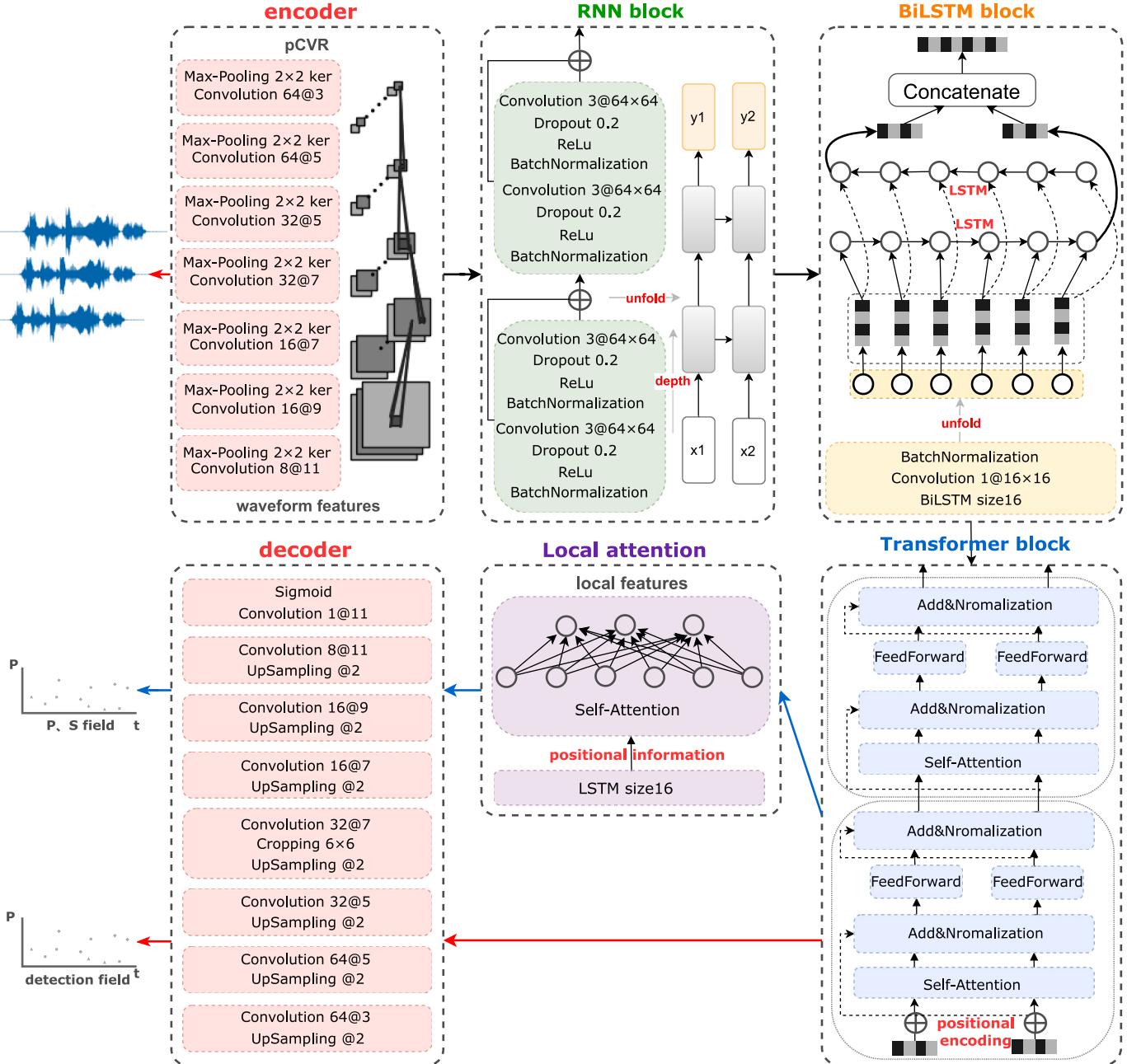


Fig. 2. The structure of the TranSeis network. The numbers indicate the number and the size of kernels, and the order of structure within the module is from bottom to top.

4. Results

4.1. Model testing results

A total of 546,949 waveforms were used to test the model. We expect that the model pays more attention to the characteristics of the changes in the wave shape itself rather than learning how to label it. To this end, manual labels were intentionally subjected to random offsets in training, ranging within a 30-s window. Fig. 5a and Fig. 5b present the test results for "standard" seismic signals. As shown in Fig. 5a and b, in addition to detecting clear seismic signals, our model also detects weak signals from the obvious fluctuations of the waveform at the sequence of 5000–7000 sample points and 10000 to 12500 sample points, respectively. The model successfully estimates the probability of signal detection at each moment. This capability is also demonstrated in signal samples with

significant noise (Fig. 5d). These results show that the model has a certain sensitivity to the waveform itself, and this ability helps avoid the disadvantages of confusing the machine when the manual label is wrong.

We extracted samples where the model's detection results showed significant discrepancies from the manual annotations (Fig. 5c and Fig. 5d). Surprisingly, the deviations were caused by errors in the manual annotations, while our model consistently maintained accurate judgments. This observation further confirms that the model has effectively learned to capture the intrinsic features of waveform variations, demonstrating its high reliability.

Moreover, we evaluated the model's robustness to noise. As shown in Fig. 5e and f, the model maintained high resolution even under substantial background noise, even when data from the third channel was insufficient.

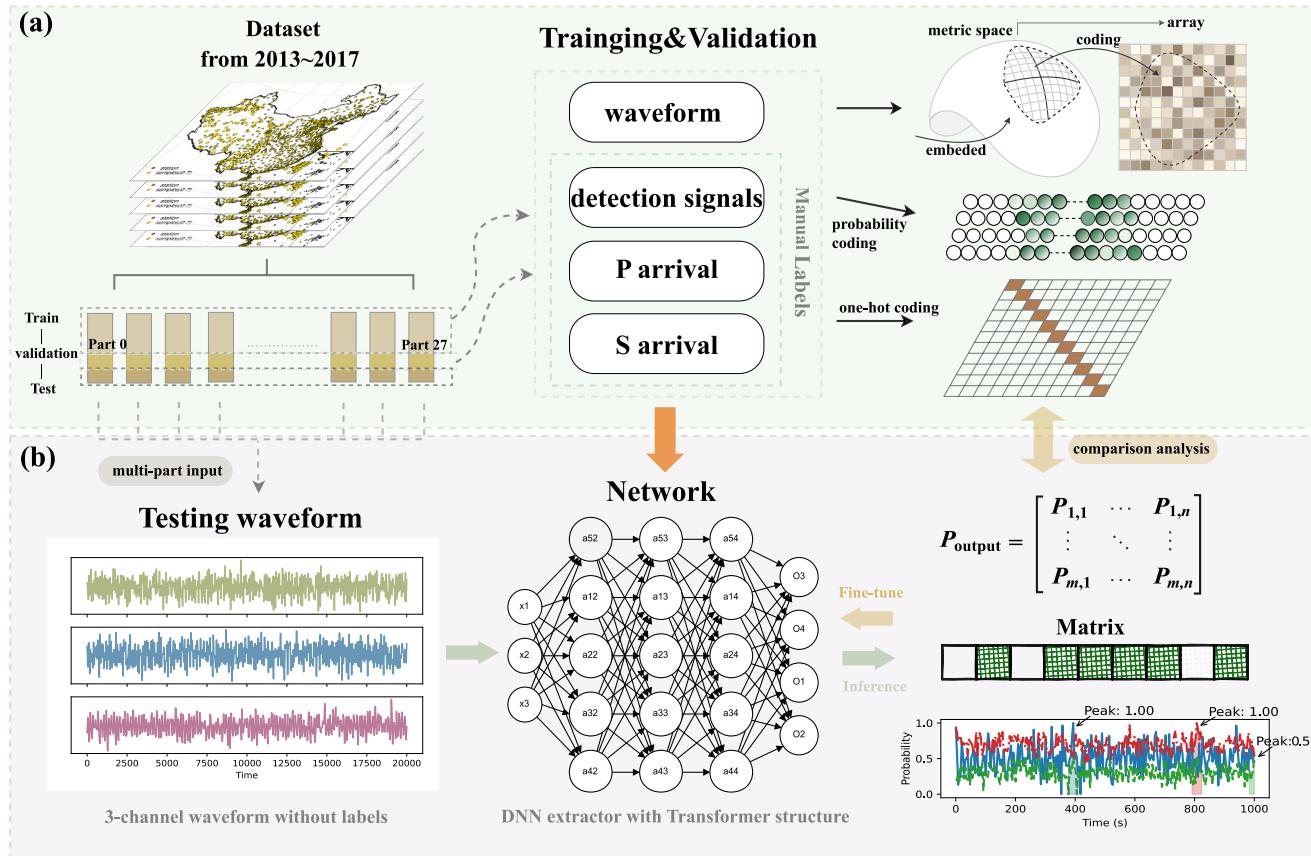


Figure 3. The main training (green box) and testing processes (purple box). (a)The dataset is divided into 28 parts, with a training, validation, and testing split of 6:2:2. Waveform data are embedded, while detection signal labels are probabilistically assigned. Specifically, sample points are randomly taken from the interval between the P-wave arrival time and the S-wave arrival time plus $c \times (S-P)$, where c is a variable parameter, with the remaining points assigned a value of zero. The labels of P and S are one-hot encoded. (b)Three-component waveforms without labels are input into the network for inference, partitioned into 28 parts. The network outputs a probability matrix and evaluation matrix used for model evaluation, fine-tuning, and selecting the points of maximum probability for seismic phase picking.

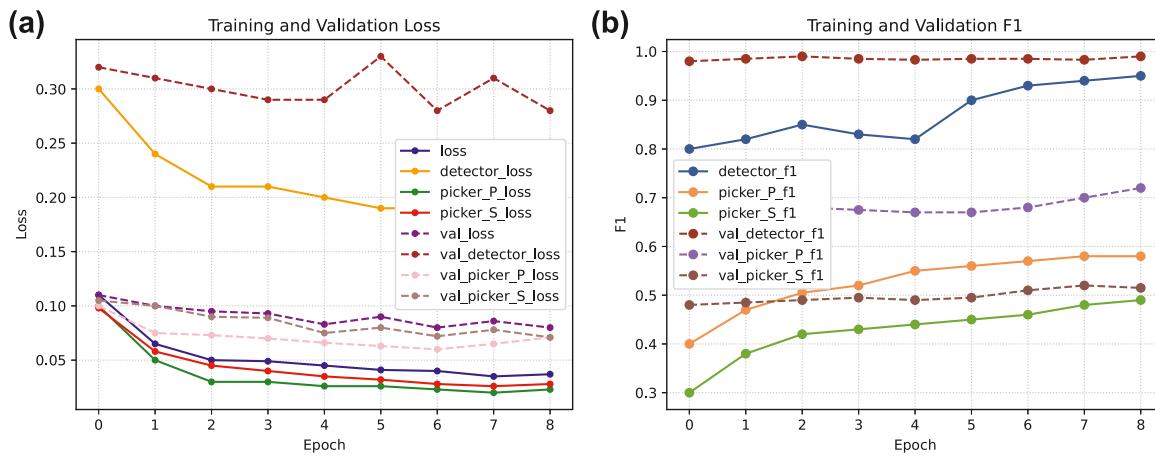


Fig. 4. Loss (a) and F1 score(b) for training and validation.

In accessing the predictive accuracy of the model, in addition to using loss and F1-score evaluation indicators, this study also incorporates the absolute distance evaluation index directly. The apparent velocity of the pseudo-waves was calculated based on the time intervals between the identified phases. With a pseudo-wave velocity of 8 km/s, an error of 1 s in the S-P measurement could potentially result in an 8 km

deviation in the epicenter distance. From this perspective, to ensure the precision of the epicenter location for near-earthquake events, we constrain the error to within 0.5 s.

We calculate the joint prediction accuracy of all samples with noise perturbations after each iteration, as shown in Fig. 6, and the average value of all accuracy across all iterations is illustrated in Table 1.

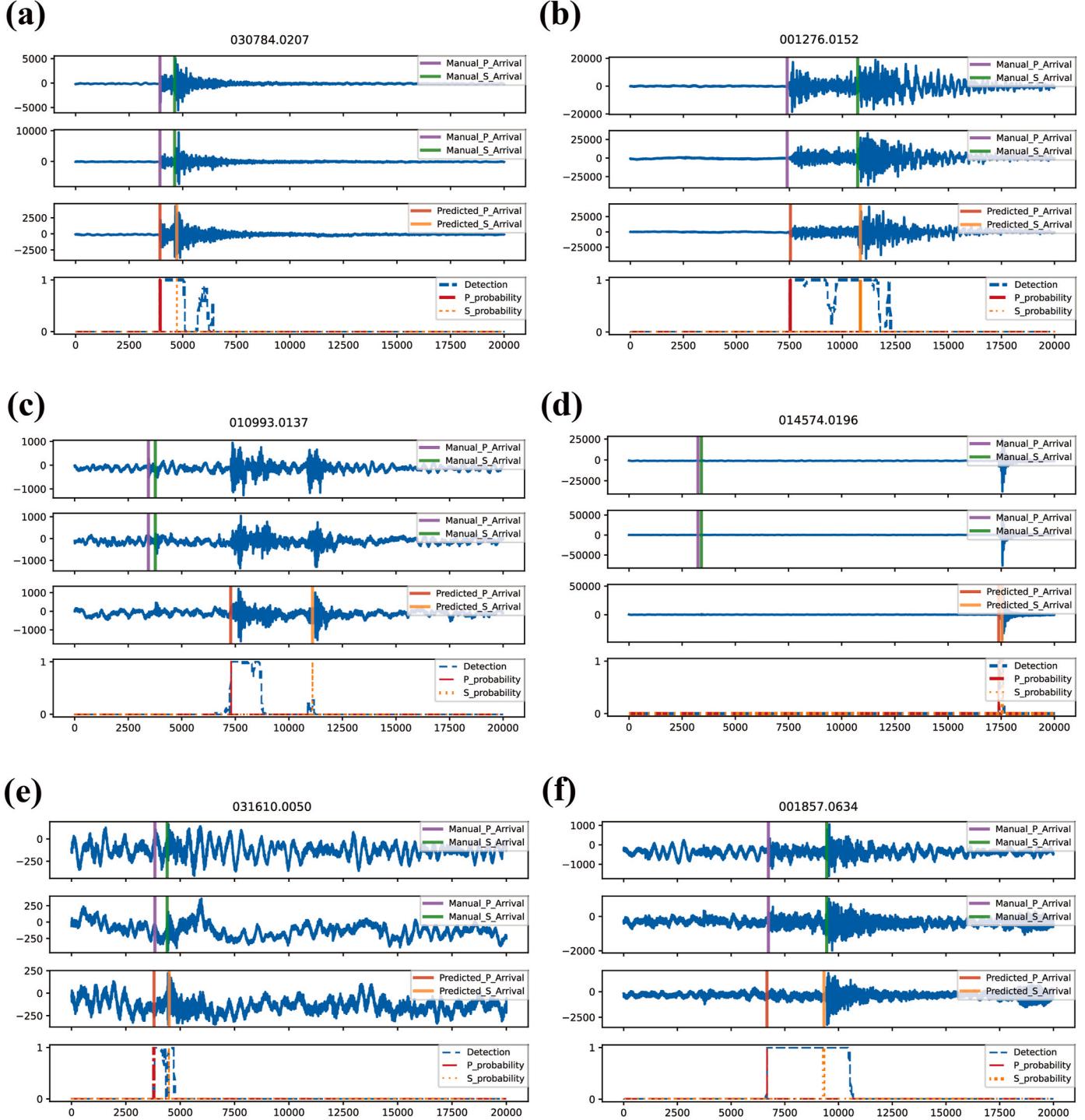


Fig. 5. Test results. (a–b) Results for "standard" seismic signals. In addition to detecting prominent seismic events, the model exhibits sensitivity to the subtle fluctuations in the waveforms, including those with significant noise. (c–d) The risks of confusion encountered by machines due to errors in manual labeling are reduced thanks to the model's stability. (e–f) The model is robust to noise.

In our experiment, to verify the positive correlation between dataset size and final prediction accuracy, we first trained and tested only one part of the dataset. The results (Table 1) show that the prediction accuracy was 92% for P-waves and 76% for S-waves. Upon expanding the training and testing datasets to encompass three parts, the prediction accuracy increased to 96% for P-waves and 88% for S-waves. Subsequently, all datasets were trained and tested. The prediction accuracy was 97% for P-waves and 87% for S-waves. From the results, it turns out the prediction accuracy of the model is positively correlated with the

size of the dataset. In contrast, the prediction accuracy of the S-wave decreases slightly. We hypothesize that this decrease may indicate potential overfitting in the S-wave picking, although the slight decrease may be merely an accidental phenomenon.

4.2. Comparison results

To further evaluate the performance of TranSeis, we compared it with two widely recognized deep-learning models on 100k test

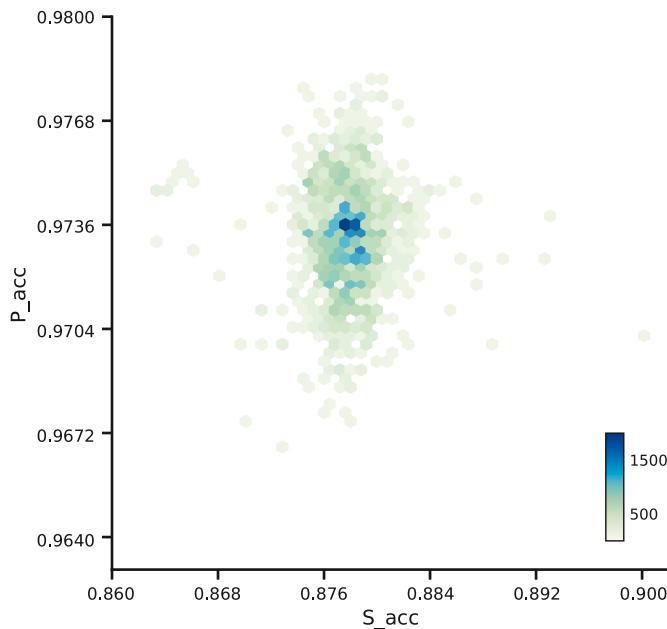


Fig. 6. The model's joint accuracy. The joint accuracy of P-S phase predictions is presented for each testing round, reflecting the model's overall performance across phases.

Table 1

Accuracy in the presence of noise perturbations at different data sizes. It turns out that the prediction accuracy of the model is positively correlated with the size of the dataset.

Data Size	Phase	
	P Accuracy	S Accuracy
1 part	0.9289	0.7652
3 parts	0.9602	0.8856
all parts	0.9733	0.8751

waveforms. These three-component waveforms were all sourced from the DiTing dataset. The deep learning models used for comparison were pre-trained on different datasets: PhaseNet was trained on the INSTANCE dataset, while EQTransformer was trained on the STEAD dataset. Inference for all three models was conducted under identical hardware and software conditions.

In practical seismic analysis, P-wave and S-wave pickups occur simultaneously. Therefore, we employ a joint P-S-wave accuracy assessment criteria. A prediction is considered a true positive if the absolute difference between the predicted and actual arrival times is less than 0.5 s. Fig. 7 visualizes the accuracy distribution for each model. These comparisons are focused on evaluating the overall performance of the models in terms of accuracy, particularly when different aspects of the signal data are processed. The green hotspots in these density comparison plots on the left side in Fig. 7 (a–c, e) represent the calculated raw accuracies. The density plots on the right side in Fig. 7(b–d, f) show the accuracy distribution after rounding the accuracy values to two decimal places and swapping the axes (i.e., swapping the axes of P and S accuracy). We carry out the experiments in three scenarios for comparison. We first calculate the accuracy without removing missed detections after introducing noisy samples into the data in Fig. 7a–b, which presents the model's noise tolerance under different noise levels. Then, in Fig. 7c–d, the accuracy of each model is recalculated after removing the missed detections for both P-wave and S-wave predictions. This comparison highlights the impact of excluding missed detections on the performance. Fig. 7e–f show accuracy, including missed detections for both P and S waves, offering an original comparison between the

models' performance under these conditions. Table 2 presents detailed comparison results of these three picking models.

The results indicate that TranSeis improves the accuracy of both P-wave and S-wave picking, with a more pronounced improvement in P-wave picking. Additionally, these results reveal several interesting phenomena.

- a) While PhaseNet demonstrates relatively low accuracy in phase picking, it exhibits an exceptionally high detection rate, with no missed detections observed in our benchmark tests. This highlights an important observation: it is sometimes easier to detect a signal than to confirm its absence.
- b) EQTransformer, in our tests, shows stronger performance in picking S-wave phases compared to P-wave phases. This is a rare and valuable capability, which may be attributed to its hierarchical attention mechanism. This mechanism likely allows the model to better adjust its focus on S-wave signals. Both Fig. 7 and Table 2 reflect this characteristic.
- c) TranSeis demonstrates remarkable stability across both noisy scenarios and cases with missing detections. In the comparison, TranSeis shows maturity and calmness, like a steady older man who is good at dealing with unexpected situations and is able to find a balance quickly. We regard TranSeis as a robust and reliable model, a trait that proves to be especially important in all practical applications.

4.3. Application to DiTing2.0 dataset

The DiTing 2.0 dataset includes phase observation reports from the China Earthquake Networks spanning from March 2020 to February 2023, containing 1,089,920 three-component waveforms from 264,298 natural seismic events in mainland China and its neighboring regions, along with 958,076 Pg. and 780,603 Sg phase arrival labels. The magnitudes of the seismic events in the dataset range from 0 to 8.2, all of which are meticulously annotated to correspond with the waveforms.

The dataset includes a unique noise dataset and consists of approximately 130,000 misidentified waveform data samples detected by artificial intelligence algorithms from continuous waveform records across the national network between June 2021 and February 2023. Given the independence between the two datasets, we are confident that the application results on DiTing 2.0 can effectively validate the model's generalization across the Chinese mainland, not only geographically but also in terms of spatiotemporal generalization. Consequently, we employed the pre-trained model TranSeis to predict the large-scale waveform dataset for nearly three years, with a portion of prediction results shown in Fig. 8 and errors shown in Fig. 9.

Fig. 8a and Fig. 8b present the prediction results for standard seismic signals across three channels: N, E, and Z. The red and orange dashed lines indicate the model's predicted arrival time of P-wave and S-wave, respectively. The fourth subplot shows the probability of signal detection and the corresponding probabilities for P- and S-wave arrivals. The model maintains its reliability in prediction, even when applied to samples exhibiting considerable variance across the three channels (Fig. 8c and Fig. 8d) or those characterized by high levels of background noise (Fig. 8e).

4.4. Application to TXED dataset

We evaluate the performance of the proposed TranSeis model using the regional artificial intelligence (AI) earthquake dataset (TXED) (Chen et al., 2024) compiled for the state of Texas. The TXED dataset is composed of earthquake signals with manually picked P- and S-wave arrival times and manually picked noise waveforms corresponding to more than 20,000 earthquake events spanning from the beginning of the Texas seismological network (TexNet) (1 January 2017) to date. The TXED dataset was sampled at 100 Hz with a waveform duration of 60 s.

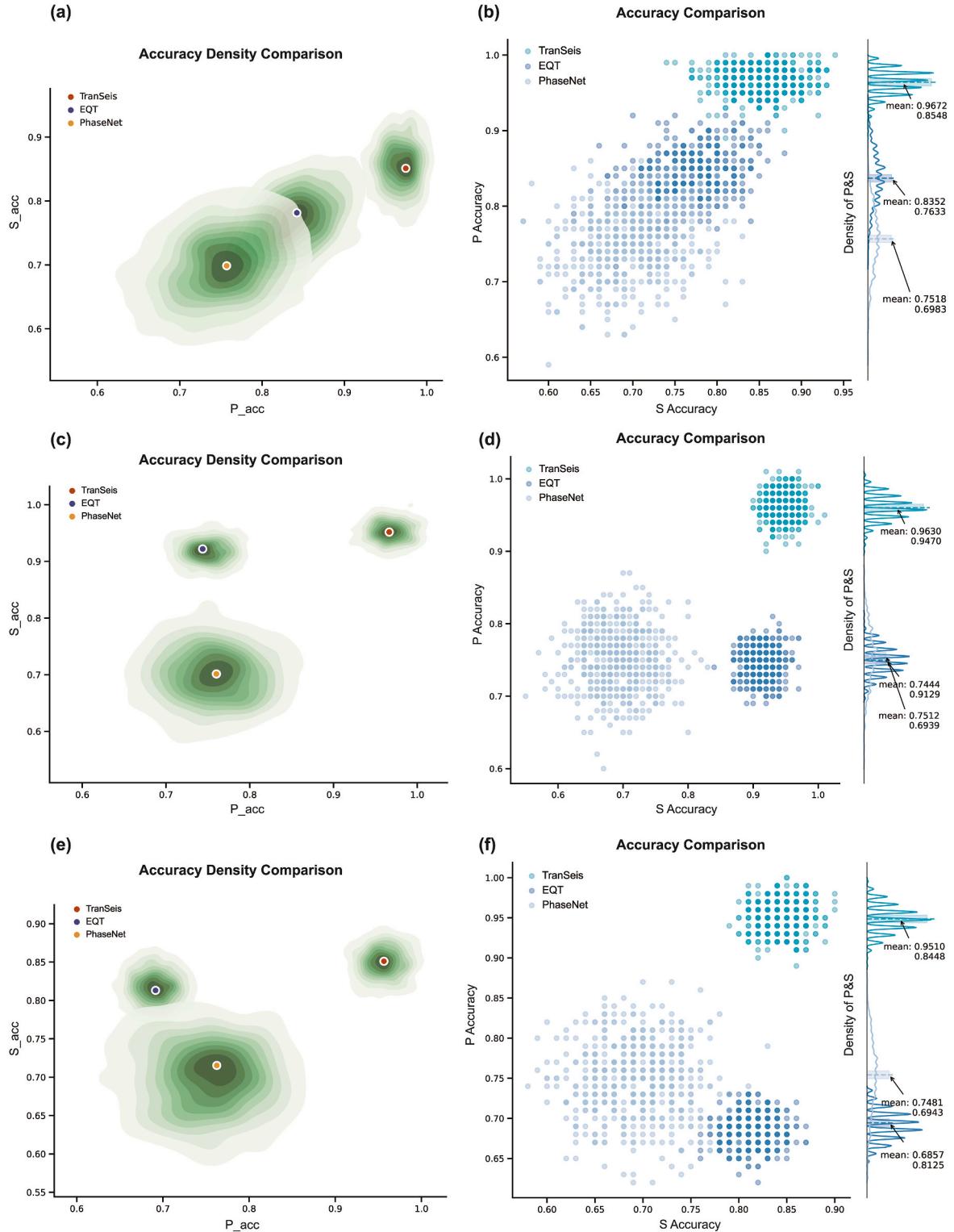


Fig. 7. Joint accuracy density comparison between three seismic phase picking models: TranSeis, EQTransformer and PhaseNet. The green contours depict the joint density distribution of the model accuracies, with higher-density regions indicated by more concentrated contours. The right plots are the density plots that show the accuracy distribution after rounding the accuracy values to two decimal places and swapping the axes. **(a-b)** The accuracy density comparison where the accuracy is calculated without removing missed detections after introducing noisy samples into the data. **(c-d)** The accuracy of each model is recalculated after removing the missed detections for both P-wave and S-wave predictions. **(e-f)** The accuracy of each model is recalculated include missed detections for both P and S waves.

As a result, the performance of our model on the TXED dataset, consisting of 312,231 seismic waveforms, is shown in Fig. 10, which illustrates the results of the generalizability test conducted on the TXED dataset, providing insights into the model's performance under various

conditions. Fig. 10(a)-(c), and (f) present low signal-to-noise ratio (SNR) waveforms, showcasing the model's ability to accurately detect seismic phases even in challenging scenarios with weak signals. Fig. 10 (b) highlights the model's robustness by demonstrating its effectiveness in

Table 2

Results comparison of the picking models.

Model	Accuracy (Exclude missing)		error				Missing	
	P	S	P MAE	P σ	S MAE	S σ	P Missing	S Missing
PhaseNet	0.7481	0.6943	0.064s	0.027s	-0.123s	0.061s	0/100000	0/100000
EQTransformer	0.6857	0.8125	0.103s	0.318s	-0.081s	0.058s	200/8100	403/8100
TranSeis	0.9510	0.8448	-0.048s	0.033s	-0.082s	0.055s	204/20000	13/20000

handling waveforms with significant variance across the three channels, further validating its noise resilience. In contrast, Fig. 10 (d) and (e) depict the model's predictions on "standard" seismic signals, where both P-wave and S-wave arrival times are identified with high precision. These results collectively underscore the model's adaptability and robustness across diverse seismic data conditions, affirming its reliability for practical applications.

In summary, the error distribution of TranSeis on this dataset can be observed in the boxplot of Fig. 8c. For P-waves, the error remains within 0.5 s, with a mean value of -0.02424s and a variance of 0.03879 s. For S-waves, although the fluctuation is slightly larger, the boxplot still indicates that the majority of errors remain within 0.5 s. Even with greater variance, the calculated mean and variance for S-wave detection are 0.2059s and 0.4976s, respectively.

4.5. Application to continuous dataset

We place great emphasis on the practical value and scientific significance of deep learning models, and we aim to investigate how our picker can assist in detecting earthquakes within seismic networks. To this end, we applied TranSeis to real-time continuous waveform data collected in this study to evaluate the model's robustness in scenarios such as network-based detection, operational use, or real-time applications.

Southwest China, particularly the Sichuan-Yunnan region, features a complex system of active faults. We selected areas along major active fault zones within this region as our test area. Specifically, Qiaojia, located at the southern end of the Zemuhe Fault Zone and the northern end of the Xiaojiang Fault Zone, lies at the intersection of these two fault systems. Within this test region, 201 short-period stations and 10 broadband stations were deployed, recording continuous waveforms over a period of 207 days from August 25, 2022, to March 19, 2023. Most signals were collected from broadband instruments, which are capable of detecting long-period signals due to their broad frequency range. Short-period instruments, on the other hand, are sensitive to signals from the surface to the upper crust. The combination of data from broadband and short-period seismic stations ensures a comprehensive record of seismic activity.

The waveforms we collected were sampled at 100 Hz and stored in 'mseed' format. First, we preprocessed the collected data, including denoising, before converting them into h5 format. Fig. 11 provides an overview of the TranSeis model's prediction capabilities on continuous data across various waveform characteristics and scenarios, showcasing its ability to generalize and maintain accuracy in diverse contexts. For these waveforms with high amplitude and clear P/S phases (Fig. 11a and b,e), the model accurately identifies both P and S arrival times, aligning closely with visible phase onsets. The clear separation of the P-phase and S-phase phases in these examples demonstrates the model's ability to perform well under favorable signal conditions. Detection probabilities in these cases show sharp and distinct peaks, reflecting the model's high confidence. Even under challenging conditions, such as low amplitude or low SNR (Fig. 11 f), the model effectively captures the subtle transitions corresponding to P-wave and S-wave arrivals. This highlights the model's robust noise-handling capabilities, as evidenced by its ability to maintain accurate predictions when waveform clarity is compromised. In waveforms where seismic phases overlap or show high variability (Fig. 11 c), the model continues to predict P and S phases with

reasonable accuracy, which confirms the model's ability to assign appropriate detection probabilities even under complex signal conditions.

In summary, across all waveforms, the TranSeis model demonstrates consistent performance to adapt to a wide range of seismic waveform characteristics, including high-SNR signals, low-amplitude noise-heavy data, and overlapping phases. This consistency underscores the model's potential for practical deployment in real-time seismic monitoring systems, even in regions with varied geological and noise conditions.

5. Discussions

5.1. Improvements and advantages

5.1.1. Improvements in model structure and efficiency

Through extensive experiments, this paper identifies an optimized architecture that not only improves accuracy but also significantly reduces computational complexity without compromising precision. We build upon the multitask framework of the EQT model, integrating signal detection and phase picking into a cohesive system. However, we employ a different architectural approach for seismic signal detection and phase picking, resulting in a more modular, compact, and efficient structure for seismic signal processing. By distributing the dataset across multiple GPUs, memory overflow issues are effectively mitigated. In contrast to previous deep learning models for seismic detection, this paper introduces a bold and direct evaluation criterion to assess the model performance. Furthermore, the emphasis on joint P-S phase accuracy, rather than isolated phase evaluation, provides a more practical and comprehensive measure of the model's real-world applicability.

A significant improvement lies in reducing the use of RNN and BiLSTM layers, retaining only two RNN layers and one BiLSTM layer. RNNs are effective for sequential data, passing hidden states between layers to capture temporal dependencies. BiLSTMs further enhance this by considering both forward and backward contexts. However, in seismic waveform data, such as P- and S-wave arrival detection, short-term and local variations are more critical. Therefore, a single BiLSTM layer is sufficient for this task, and adding more layers would only increase computational overhead without significant performance gain. By simplifying the network, the model effectively captures essential temporal features while improving training speed and generalization.

Furthermore, the removal of the LSTM layer before the Transformer module and the reduction of hidden units in the decoder further decrease the model's size and complexity. These adjustments improve computational efficiency without compromising accuracy.

Experimental results confirm that the simplified architecture maintains high accuracy on seismic datasets and performs robustly in noisy conditions. Reducing RNN and BiLSTM layers not only avoids overfitting but also strikes a balance between accuracy and computational efficiency. It demonstrates that a carefully streamlined model can outperform more complex architectures, offering valuable insights for neural network applications in seismology.

5.1.2. New evaluation metric

Traditional evaluation metrics such as F1-score, recall, and precision are widely used in classification tasks, focusing on the model's ability to correctly identify positive and negative samples, especially in imbalanced datasets. However, in the context of seismic phase picking, where

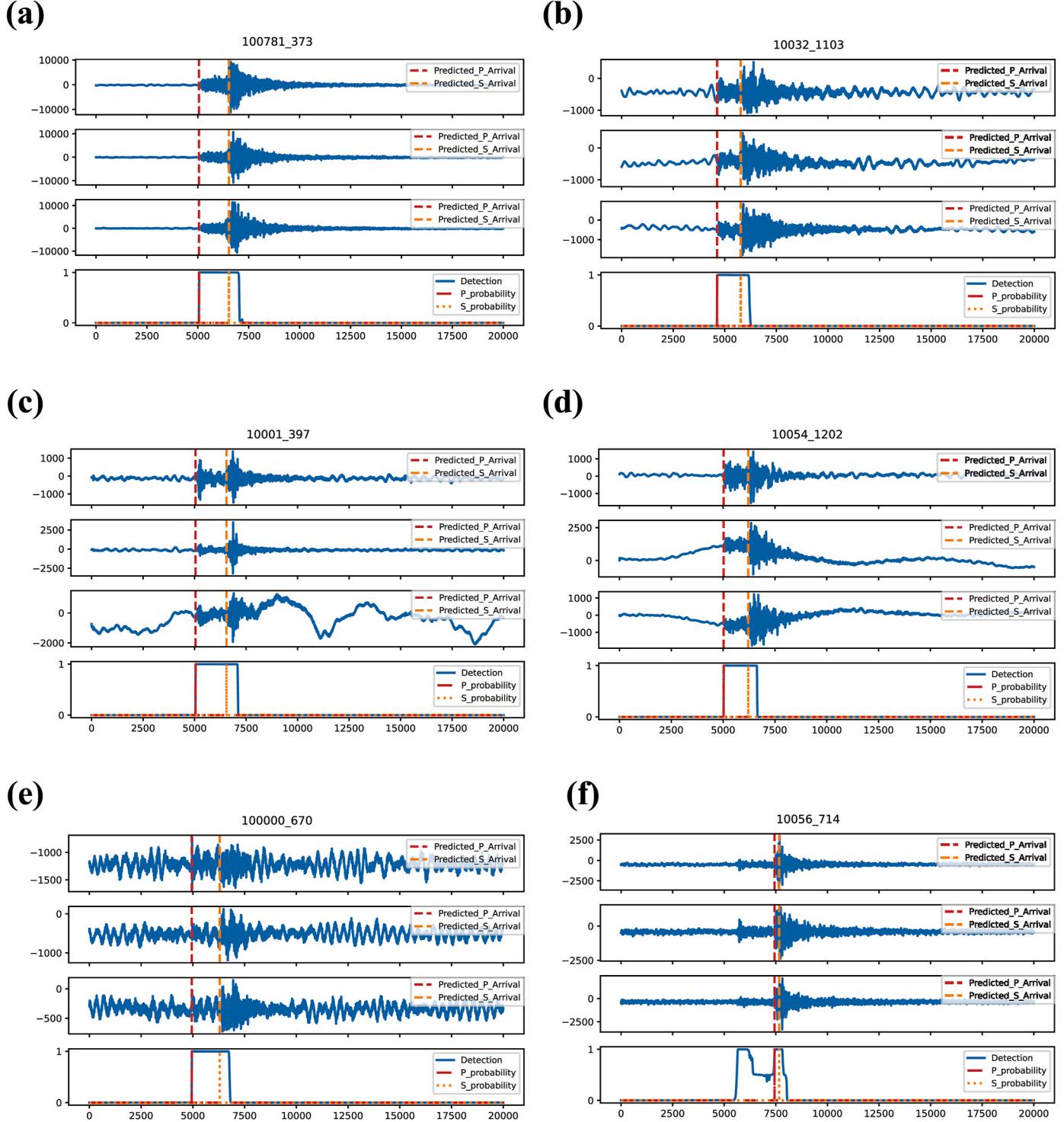


Fig. 8. Generalizability test results on DiTing2.0 dataset. The red dashed line indicates the P-wave arrival predicted by the model, while the orange dashed line represents the predicted S-wave arrival. The fourth subplot displays the predicted probabilities of signal detection and P and S arrivals. (a-b) The model's predictions for "standard" seismic signals. (c-e) The model's performance on samples with high variance across the three channels further validates the model's robustness to noise. (f) The model predicts the possible arrivals from the subtle waveform variations.

The goal is to predict the arrival times of seismic waves with high precision to support tasks such as hypocenter location and magnitude estimation; these traditional metrics can be inadequate. A slight time error may significantly impact the accuracy of subsequent analyses, such as earthquake location and magnitude determination, even if the model's overall performance seems satisfactory. This localized error may not be fully reflected in metrics like F1-score or recall, as these

metrics focus more on overall classification performance. Therefore, a more direct metric is needed to evaluate the model's ability to predict the arrival times of seismic phases accurately. In traditional seismic phase picking tasks, the accuracy of P-wave and S-wave predictions is typically evaluated independently. However, in practical seismic analysis, the picking of P- and S-waves occurs simultaneously. Therefore, this paper adopts a joint P-S accuracy evaluation method rather than

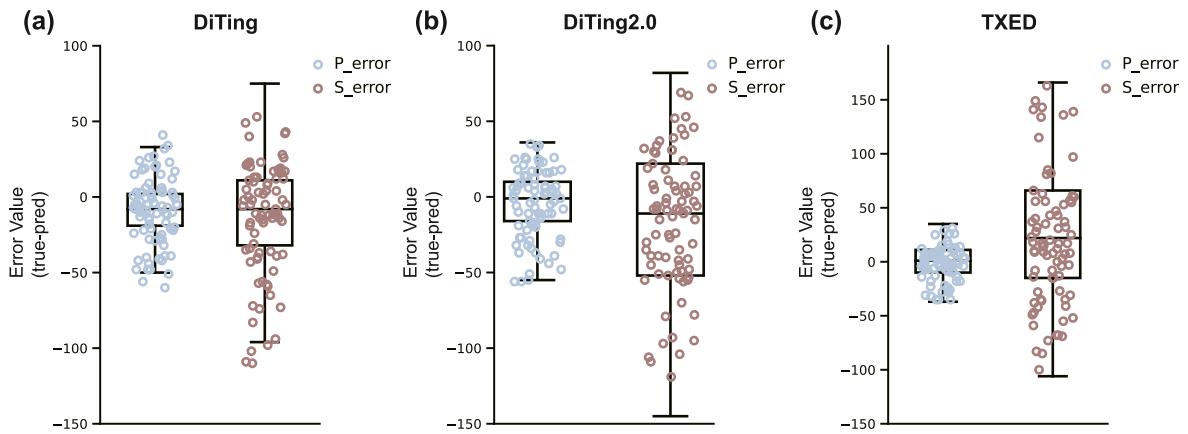


Fig. 9. The seismic phase picking errors (the difference between the actual and predicted arrival times). P picking error (blue box) remains consistently within 0.5 s with low variance. S picking error shows a slightly larger variance, and the majority of errors are also within 0.5 s. (a) Testing errors on DiTing dataset, (b) DiTing2.0 dataset and (c) TXED dataset.

assessing the phases in isolation. This approach more accurately reflects the model's ability to predict both P- and S-waves within the same seismic event. By focusing on joint accuracy, the evaluation aligns more closely with real-world seismological studies, offering a more comprehensive and realistic assessment of the model's performance.

This work introduces absolute error as the core evaluation metric to address this issue. Absolute error directly measures the time difference between the predicted and true seismic phase arrivals, providing a more precise reflection of the model's ability to meet the core requirements of the task. By calculating the time deviation for each prediction, the accuracy of the model becomes clearer and more intuitively understood. This error metric is particularly significant for high-precision tasks, where even minor time discrepancies may lead to considerable consequences.

While absolute error is introduced as the central metric, traditional metrics such as F1-score, precision, and recall are not entirely discarded. They are still useful for evaluating the overall classification performance of the model. By combining these traditional metrics with absolute error, the paper presents a more comprehensive evaluation approach—one that captures both the model's general performance and its accuracy in predicting key time points.

Extensive experiments validate the effectiveness of absolute error, demonstrating its superiority over traditional classification metrics in the context of seismic phase picking. When compared to other existing models, the proposed model achieves lower absolute error, particularly in high-noise environments and with complex waveforms, while maintaining high accuracy. The introduction of absolute error as a new evaluation metric not only provides a more reasonable and precise assessment of the model's performance in seismic phase picking but also highlights the need for the broader adoption of such metrics in future studies within the field.

5.2. Main contribution

This work makes significant contributions to addressing two key scientific issues related to earthquake prediction and precursor information extraction. The first challenge we tackle is the recognition and enhancement of weak seismic signals, which are critical for identifying earthquake precursors. By leveraging advanced deep learning techniques, particularly the TranSeis model based on the Transformer architecture, we achieve improved accuracy and efficiency in detecting weak signals in seismic monitoring that are often undetectable using traditional methods. This capability to identify subtle seismic signals within noisy and complex data significantly enhances our ability to predict major seismic events. The second scientific challenge we address is the extraction of reliable precursor information from seismic

waveforms using deep learning models. By designing TranSeis, we provide a powerful tool for detecting seismic signals with high accuracy. The model is specifically tailored to the unique seismic characteristics of mainland China, where the complex topography and geological structures render existing earthquake detectors less effective. TranSeis's deep neural network approach facilitates the accurate extraction of key seismic features. By applying state-of-the-art artificial intelligence to this problem, we improve the reliability of earthquake prediction systems and contribute to a more advanced understanding of seismic activity.

6. Conclusions and future perspectives

6.1. Conclusions

We have developed a high-precision deep learning model based on the Transformer architecture named TranSeis, which demonstrates exceptional capabilities in earthquake detection and P- and S-wave phase picking using seismic data from mainland China. Even under complex and high-noise conditions, its multi-task framework and parallel processing capabilities enable accurate identification of P- and S-phase identification while maintaining computational efficiency, making it highly practical for real-time applications in seismic networks. In benchmark tests against state-of-the-art models such as PhaseNet and EQTransformer, TranSeis exhibits more robust and reliable performance, ensuring higher accuracy. Furthermore, it demonstrates strong generalization capabilities across diverse datasets (e.g., DiTing2.0, TXED) and continuous waveform data collected from tectonically complex regions, such as the Sichuan-Yunnan area in Southwest China.

The application of TranSeis to continuous seismic data highlights the transformative potential of data-driven methodologies, facilitating the direct exploration of earthquake characteristics and anomalies from waveform data. This approach not only contributes to the development of a robust framework for precursor signal identification but also enhances earthquake prediction efforts, effectively addressing the challenges posed by the growing volume of seismic data.

Additionally, during benchmark comparisons with PhaseNet and EQTransformer, we identified the unique strengths of each model. PhaseNet demonstrates exceptional performance in minimizing missed detections, while EQTransformer excels in S-wave phase picking, which is particularly noteworthy. TranSeis effectively balances these two strengths, achieving both high accuracy and versatility. This adaptability makes it suitable for a variety of practical applications, enabling users to select the appropriate model depending on specific scenarios. In some instances, the integration of these models can be employed flexibly to optimize performance across different tasks.

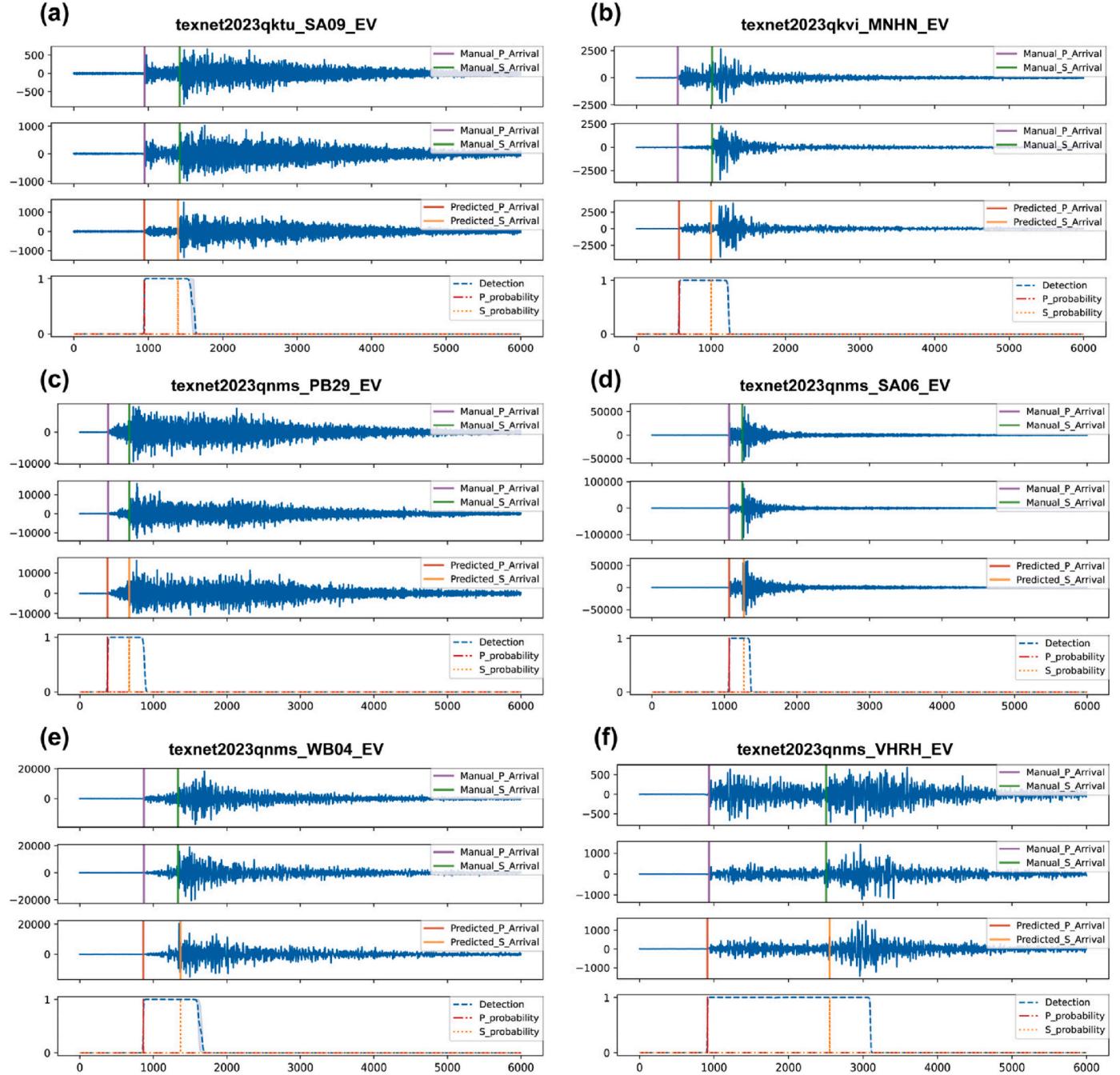


Fig. 10. Generalizability test results on TXED dataset. (a) and (c) and (f) Low SNR waveforms. (b) The model's performance on samples with high variance across the three channels further validates the model's robustness to noise. (d) and (e) The model's predictions for "standard" seismic signals.

6.2. Future perspectives

The successful application of deep learning models in seismological research demonstrates that these advanced models may uncover nonlinear features and complex temporal sequences within seismic signals. This capability has the potential to highlight the limitations of current theories and open new avenues for seismological investigation. Future research will necessitate a comprehensive analysis of model results and an exploration of interpretability at the theoretical level to achieve a deeper understanding of seismic phenomena and improve the accuracy of earthquake predictions.

The model's ability to utilize waveform information for predicting seismic signals and phases demonstrates its capacity to learn from waveforms. We believe that this learning capability can be further

leveraged. In the next phase of our research, we aim to enhance the model's generalization ability by refining the algorithms and training it on a more diverse array of datasets. This approach will enable the model to adapt to various regions and types of seismic activity and phase pickup. Additionally, we plan to integrate the model into real-time seismic monitoring networks to facilitate rapid earthquake detection and early warning systems. This paper introduces new evaluation criteria for seismological detection models. It is important to establish harmonized, task-specific evaluation principles to support future comparisons and advancements in the field. This paper serves as a reference for addressing this need. Furthermore, the multitasking parallel model structure exhibits significant potential, raising the question of whether the model can also perform additional tasks, such as epicenter determination and polarity prediction.

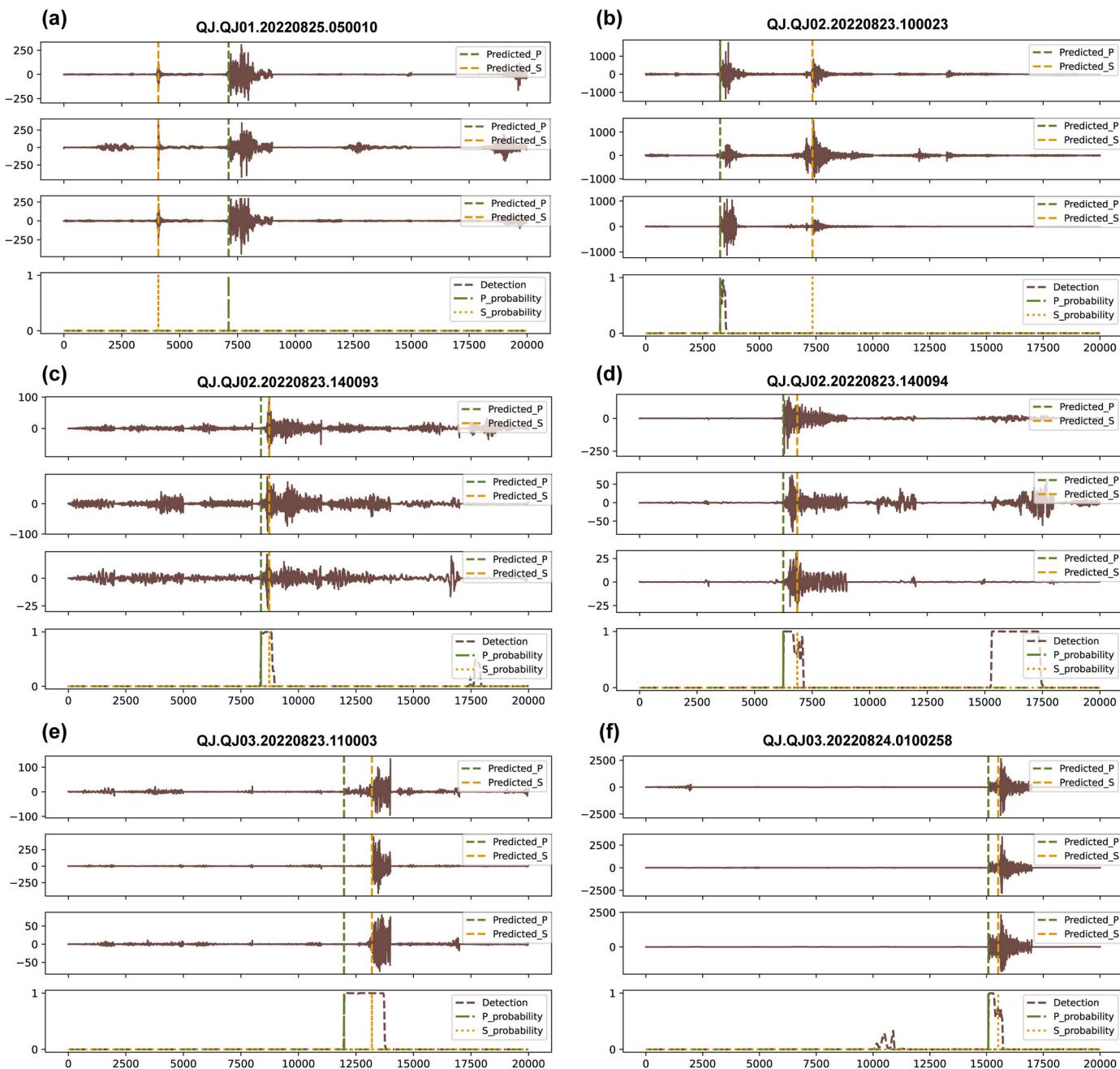


Fig. 11. Results of continuous waveforms. (a, b, e) High amplitude and clear P/S Phases. (c, d) Seismic phases overlap or show high variability. (f) Low SNR waveform.

CRediT authorship contribution statement

Yuxin Zhou: Writing – original draft, Validation, Software, Methodology, Formal analysis. **Huai Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Shi Chen:** Data curation. **Zheng Yuan:** Software. **Chuanqi Tan:** Software. **Fei Huang:** Writing – review & editing, Resources, Project administration. **Yicun Guo:** Funding acquisition. **Yaolin Shi:** Supervision.

Code availability section

The DiTing (Zhao et al., 2023) and DiTing2.0 datasets are shared online through the website of the China Earthquake Data Center (<https://data.earthquake.cn>), allowing users to retrieve the metadata information of the dataset online. The TXED dataset is available at <https://github.com/chenyk1990/txed>.

The source codes and sample data are available for downloading at the link: <https://github.com/Chouyuin/TranSeis>.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Key R&D Program (2023YFC3007303), the Joint Funds of the National Natural Science Foundation of China (No. U2239205), and the Fundamental Research

Funds for the Central Universities. Alibaba Group supported this work through the Alibaba Research Intern Program. We greatly appreciate the researchers at the China Earthquake Networks Center (National Earthquake Science Data Center) for providing the large-scale artificial intelligence seismology training dataset ‘DiTing’ and ‘DiTing2.0’ datasets. Thanks to Dr. S. Mostafa Mousavi of Stanford University for providing the related open-source code.

Data availability

Data will be made available on request.

References

- Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* 68, 1521–1532.
- Baer, M., Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events. *Bull. Seismol. Soc. Am.* 77, 1437–1445.
- Chan, W., Jaitly, N., Le, Q., Vinyals, O., 2016. Listen, attend and spell: a neural network for large vocabulary conversational speech recognition. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4960–4964.
- Chen, Y., Savvaidis, A., Saad, O.M., Dine Huang, G.C., Siervo, D., O’Sullivan, V., McCabe, C., Uku, B., Fleck, P., Burke, G., Alvarez, N.L., Domino, J., Grigoratos, I., 2024. TXED: the Texas earthquake dataset for AI. *Seismol Res. Lett.* 95, 2013–2022.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
- Huang, Z., Xu, W., Yu, K., 2015. Bidirectional LSTM-CRF Models for Sequence Tagging.
- Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.
- Jiang, C., Fang, L., Fan, L., Li, B., 2021. Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthq. Sci.* 34, 425–435.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *CoRR* abs/1412.6980.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Lin Xuekai, X.C., 2022. Deep-learning-empowered earthquake catalog building: comparison and evaluation of PhaseNet and EqT models. *Geomatics Inf. Sci. Wuhan Univ.* 47, 855–865.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation.
- Luo, D., Kang, H., Long, J., Zhang, J., Liu, X., Quan, T., 2023. GDN: guided down-sampling network for real-time semantic segmentation. *Neurocomputing* 520, 205–215.
- Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., Lauciani, V., 2021. Instance – the Italian seismic dataset for machine learning. *Earth Syst. Sci. Data* 13, 5509–5544.
- Ming, Z., Shi, C., LiHua, F., David, A.Y., 2019. Earthquake phase arrival auto-picking based on U-shaped convolutional neural network. *Chinese Journal of Geophysics (in Chinese)* VL - 62, 3034–3042.
- 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. *Nat. Commun.* 11, 3952.
- Mousavi, S.M., Sheng, Y., Zhu, W., Beroza, G.C., 2019. STanford EAarthquake dataset (STEAD): a global data set of seismic signals for AI. *IEEE Access* 7, 179464–179476.
- Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. *Sci. Adv.* 4, e1700578.
- Rajbhandari, S., Rasley, J., Ruwase, O., He, Y., 2020. ZeRO: Memory Optimizations toward Training Trillion Parameter Models.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. *ArXiv* abs/1505, 04597.
- Ross, Z.E., Meier, M.A., Hauksson, E., Heaton, T.H., 2018. Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am.* 108, 2894–2901.
- Ross, Z.E., White, M.C., Vernon, F.L., Ben-Zion, Y., 2016. An improved algorithm for real-time S-wave picking with application to the (augmented) ANZA network in southern California. *Bull. Seismol. Soc. Am.* 106, 2013–2022.
- Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., Catanzaro, B., 2019. Megatron-LM: training multi-billion parameter language models using model parallelism. *ArXiv* abs/1909.08053.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition *arXiv* 1409.1556.
- Sperber, M., Niehues, J., Neubig, G., Stüker, S., Waibel, A.H., 2018. Self-Attentional Acoustic Models. *Interspeech*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull. Seismol. Soc. Am.* 88, 95–106.
- Zhao, M., Xiao, Z., Chen, S., Fang, L., 2023. DiTing: a large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology. *Earthq. Sci.* 36, 84–94.
- Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B., 2016. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification.
- Zhu, W., Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival time picking method. *Geophys. J. Int.* 216, 261–273.