ARTICLE

OPEN

Check for updates

An all-in-one seismic phase picking, location, and association network for multi-task multi-station earthquake monitoring

Xu Si [1], Xinming Wu [1 ✉], Zefeng Li [1 ✉], Shenghou Wang[2] & Jun Zhu [1]

Earthquake monitoring is vital for understanding the physics of earthquakes and assessing seismic hazards. A standard monitoring workflow includes the interrelated and inter-dependent tasks of phase picking, association, and location. Although deep learning methods have been successfully applied to earthquake monitoring, they mostly address the tasks separately and ignore the geographic relationships among stations. Here, we propose a graph neural network that operates directly on multi-station seismic data and achieves simultaneous phase picking, association, and location. Particularly, the inter-station and inter-task physical relationships are informed in the network architecture to promote accuracy, inter-pretability, and physical consistency among cross-station and cross-task predictions. When applied to data from the Ridgecrest region and Japan, this method showed superior performance over previous deep learning-based phase-picking and localization methods. Overall, our study provides a prototype self-consistent all-in-one system of simultaneous seismic phase picking, association, and location, which has the potential for next-generation auto-mated earthquake monitoring.

---

[1] School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China. [2] Geophysical Research Institute, China Oilfield Services Limited, Tianjin, China. ✉email: xinmwu@ustc.edu.cn; zefengli@ustc.edu.cn

E arthquake monitoring is one of the most fundamental operations in seismology. A standard earthquake monitoring workflow involves a series of steps to detect and characterize earthquakes, including phase picking, association, and event location[1–3]. Phase picking, a conceptually simple task that is akin to detection problems in computer vision, has recently been improved through deep learning[3–18], where convolutional neural networks (CNNs)[19] are typically used. After the phase picking, traditional[20–23] and deep-learning-based[24–26] phase association algorithms have been used to link seismic phases at multiple stations from the same events. Finally, location algorithms[27] utilize the associated phases to obtain the earthquake hypocenters, although some deep-learning-based methods directly process raw data to locate earthquakes[28–34].

These three tasks (phase picking, association, and location) are closely interdependent. The accuracy of multi-station phase picking affects the accuracy of association and location. Conversely, association and location impose constraints on multi-station phase picking. Additionally, phase picking with multi-station data can further utilize the geographic relationships and waveform similarities among multiple stations. To achieve more efficient and accurate earthquake monitoring, a suitable earthquake monitoring workflow should impose inter-task and inter-station constraints and preferably perform all three tasks simultaneously at all stations. However, most existing earthquake monitoring methods perform phase picking, association, and earthquake location separately. In addition, most of the current phase-picking methods process seismic data on a station-by-station basis. While some recent graph-based approaches[35–39] have demonstrated the ability to handle irregularly spaced stations for phase association and event location, it remains a challenging task to develop a method that effectively leverages inter-task and inter-station constraints, and ideally performs all three tasks simultaneously.

Here, we propose an all-in-one earthquake monitoring system called seismic Phase picking, Location, and Association Network (PLAN) that achieves for the first time the simultaneous implementation of the three tasks with multi-station data and inter-task constraints. PLAN consists of four interdependent neural network modules. Specifically, the first module of waveform feature extraction utilizes an encorder-decoder architecture to extract relevant features from multi-station seismic data. The second module of earthquake location encodes station locations (i.e., longitude, latitude, and elevation) and merges them with waveform features from the first module to predict the earthquake depth and epicentral distance for each station. The third module of phase association utilizes the predicted earthquake location information to estimate the time shifts required to align multi-station waveform features. Finally, the fourth module of phase picking aggregates the aligned features for simultaneous multi-station phase picking. We applied PLAN in the Ridgecrest and Japan regions and compared its efficiency and accuracy with that of state-of-the-art phase-picking and event location methods, demonstrating the merits of inter-station and inter-task constraints for accurate earthquake monitoring.

Results

Multi-station multi-task PLAN. The proposed multi-station multi-task PLAN (Fig. 1) employs a Graph neural network (GNN)[40] as the backbone to integrate the four functional modules of waveform feature extraction, earthquake location, multi-station association, and a physics-informed multi-station phase picking (further details are provided in the Methods section). Compared with CNN, GNN is naturally suited for handling seismic data acquired from irregularly spaced stations[38].

For the GNN in PLAN, the graph nodes and the feature vectors are represented by the seismic stations and the corresponding information (i.e., locations and seismograms), respectively. All nodes are linked together and the linking weights are learnt during training to infer the relationships among the stations. We construct the GNN layers with TransformGConvs[41], which are designed based on an attention mechanism[42] to learn the dynamically linking weights among different stations. (Details about TransformGConvs are provided in the "Methods" section). In addition, the graph nodes are not fixed so that the GNN could be adapted to variations in the station number and location.

Raw three-component seismic signals are feature vectors of the graph nodes. The front-end waveform feature extraction module, constructed as an encoder-decoder CNN and shared among the nodes, extracts their corresponding key features. The station feature extraction block, constructed as two MLPs and shared among the nodes, extract geographic features from the normalized input longitudes, latitudes, and elevations of the stations. The earthquake location module then concatenates the extracted waveform and geographic features and employs multiple TransformGConvs to aggregate these features from multiple nodes to predict the event depth and station-event offset. The predicted offsets and depth are further used to determine the event location by triangulation[43]. Instead of predicting the hypocentral location, we predict the station-event offsets and the depth, and feed them into the followed multi-station association module to estimate to estimate the time shifts needed to align the P-wave and S-wave arrivals.

The multi-station association module plays a key role in bridging the tasks of earthquake location and multi-station phase picking and introduces physical constraints between the two tasks. Prior to aggregating the waveform features from different stations for multi-station phase picking, the features corresponding to the same earthquake are required to be initially aligned or associated; otherwise, aggregation of unaligned features could mutually interfere and ultimately degrade the picking performance. The multi-station phase-picking module includes a non-trainable physical layer, implemented with the Pytorch[44] roll function, to shift and align the waveform features (from the decoder of the waveform feature extraction module) using the time shifts. Subsequently, multiple TransformGConvs in the phase-picking module aggregate the aligned waveform features to enhance the phase-picking features in the aligned space. Eventually, another physical layer unshifts the aggregated features back to the original space, followed by two convolutional layers to obtain the P/S-wave picks at all the stations.

Three regression loss functions are defined for the three modules corresponding three tasks of phase picking, association, and earthquake location and then combined to jointly train the entire network. Because all the modules are interconnected within the entire network, the training process finds an optimal network that could perform all the tasks both accurately and consistently. Moreover, after training, the multi-station association module could be detached from the network and utilized to calculate the S-P differential travel time with inputs: offsets and event depth. Further details on this module are provided in the section titled "Multi-station association module".

Data preparation. We tested the proposed PLAN in two regions of Ridgecrest and Japan. For the Ridgecrest region (Fig. 2a), seismic recordings from 16 California Integrated Seismic Network stations within an epicentral distance of <80 km were collected from 1 January 2014, to 31 December 2021, for a total of more than 71,000 M > −0.5 earthquakes. The data for Japan (Fig. 3a) included M > 2 earthquakes that occurred between
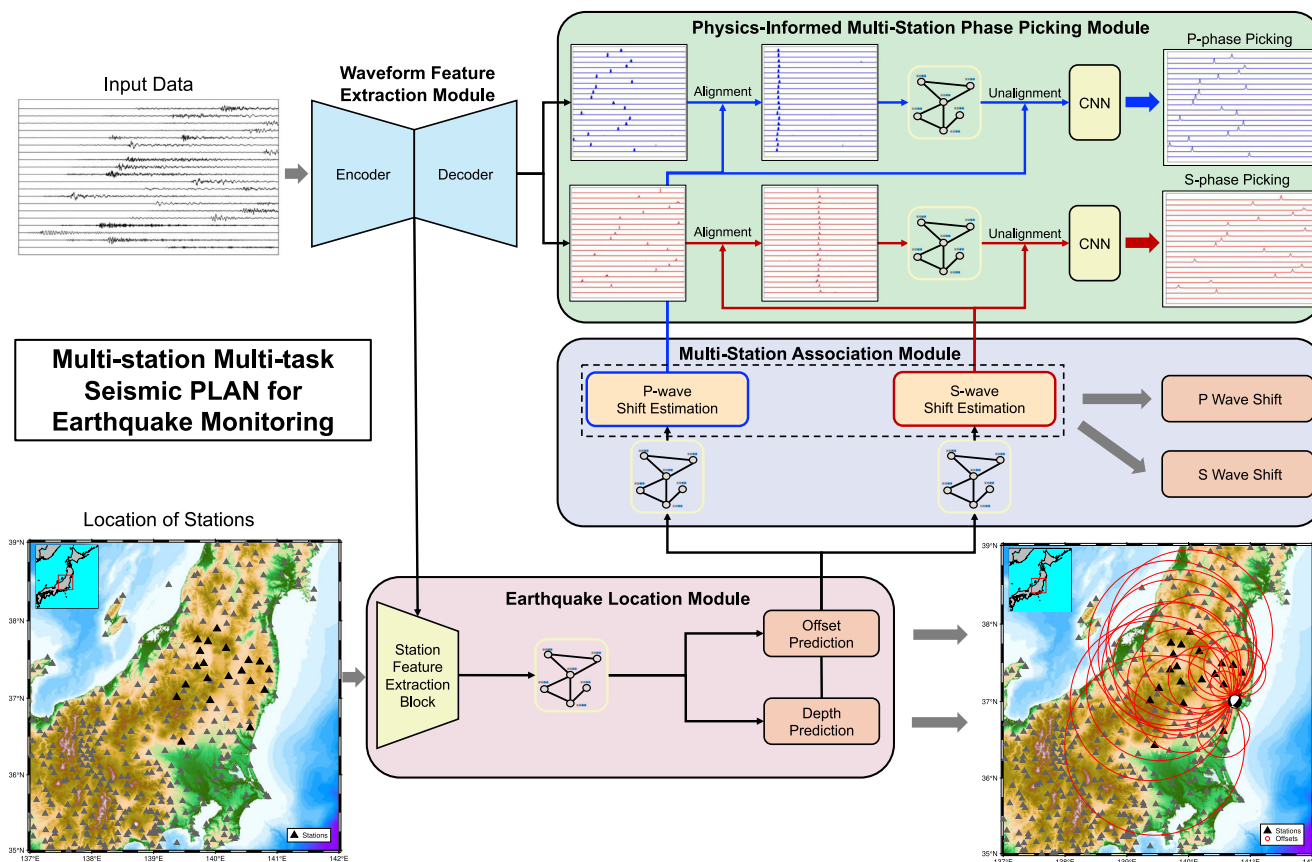
Fig. 1 The flowchart of the proposed multi-task and multi-station PLAN for earthquake monitoring. The input data for the model comprise seismic waveforms recorded by multiple stations and the locations of these stations. PLAN consists of four sub-modules: waveform feature extraction encoder and decoder, an earthquake location module, a multi-station association module and a physics-informed multi-station phase-picking module. All of these sub-modules are optimized simultaneously and constrained by each other during training to improve performance in earthquake detection, association, and location.

January 1, 2011, and December 31, 2011, including the Mw 9.1 Tohoku sequence. We collected the 3-component High Sensitivity Seismograph Network (NIED Hi-net)[45,46] data from over 35,000 events. Subsequently, the data were randomly divided into training, validation, and test sets (85%, 5%, and 10%, respectively) in both regions.

The number of stations corresponding to each event in the training samples varied, and the trained network can flexibly handle situations where the number of stations changes in actual data. Further, the distributions of the number of stations per event in the training and test sets were balanced. The results for the test sets in the two regions are presented in Figs. 4 and 5, respectively. To accommodate different range scales in the two study regions, we used different window lengths in two regions (30.72 s for Ridgecrest and 61.44 s for Japan) with the same sampling frequency (100 Hz).

To ensure a fair comparison of PLAN with the existing phase-picking methods, we followed the same data preprocessing procedures used in previous studies[6,12]:(1) normalizing the data by removing the mean and dividing by the standard deviation;(2) using a Gaussian-shaped target function as training labels for the P/S-phase arrival times. Thus, the probability vector of P/S wave is the sum of a zero vector and a Gaussian window (0.4 s), with the center of the window fixed at the P/S wave arrival time.

Application to seismicity in Ridgecrest region. We compared the performance of PLAN with that of other established deep learning methods for phase picking (PhaseNet[6] and

EQTransformer[12]) and location (Aggregated-GNN[36]). All of the methods were retrained on the same training set and evaluated on a common test set. As shown in the Ridgecrest application (Fig. 2b, c), the performance of PLAN in phase picking was superior to that of the other two deep learning-based methods. Specifically, the residual distribution of the P-wave picks for PLAN was more concentrated than that of the other methods, indicating a higher overall accuracy. For S-wave picks, PLAN performed significantly better than EQTransformer because the distribution of PLAN was narrower whereas the difference in performance between PLAN and PhaseNet was relatively minor.

In terms of localization, our method (PLAN) outperformed Aggregated-GNN (Fig. 2d–e and Table 1). The distribution of PLAN was notably more concentrated than that of Aggregated-GNN, particularly in terms of offset prediction. To further demonstrate the effectiveness of TransformGConv, we replaced all the TransformGConv layers in Supplementary Fig. 1 with GCN[40], SAGE[47], and GATv2[48], respectively. Among the various methods compared, PLAN yielded the lowest offset residual, with an average error of 1.09 km and a standard deviation of 1.41 km. Furthermore, PLAN also outperformed Aggregated-GNN in terms of depth localization, regardless of whether it was based on GCN, GATv2, or TransformGConv. These results demonstrated the superiority of the proposed PLAN in location estimations.

Furthermore, we used three metrics of mPrecision, mRecall, and mF1 (described in the Methods section), to quantitatively evaluate the performance of the five methods (Table 2). In five of
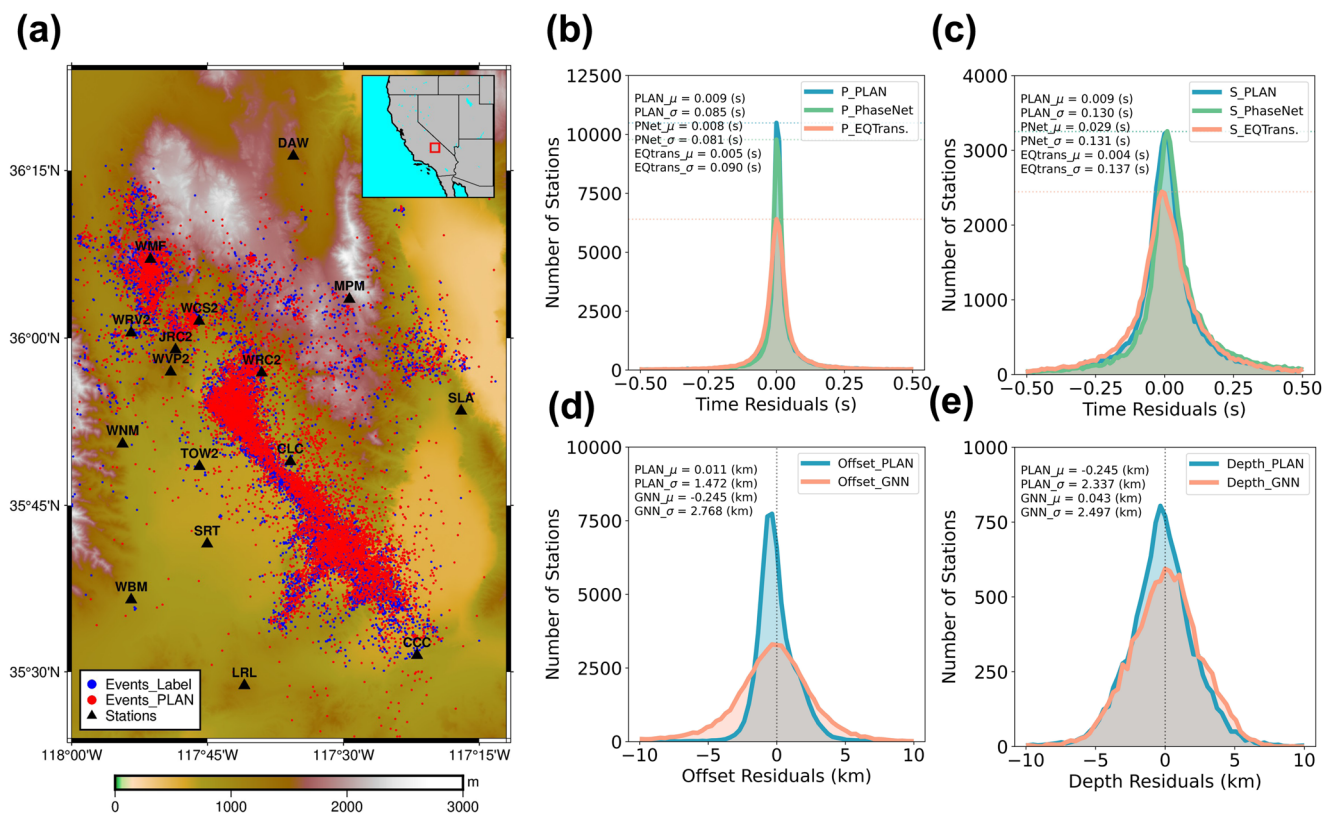
Fig. 2 Distributions of phase picking and location residuals in Ridgecrest region. a distribution of 16 stations (black triangles) and event locations of the test dataset (blue circles) used in our study. The red points represent the earthquake locations predicted by PLAN. **b** and **c** are the results of P-wave and S-wave arrival time residual, respectively. The blue, green, and orange lines in **b** and **c** represent the arrival time residuals for PLAN, PhaseNet, and EQTransform, respectively. The proposed method yields the most accurate results in P/S-wave picking. **d** and **e** represent the offset and depth residuals between model predictions and Southern California Seismic Network (SCSN) catalog of the located events. Regardless of the offset or depth, the residual distribution of PLAN (blue line) is more concentrated at zero than that of Aggregated-GNN (orange line).

the six metric scores for the P-wave and S-wave picking results, our attention mechanism-based GNN method outperformed the other methods. The only exception was the mPrecision metric of P-wave picking, where the EQTransformer showed slightly higher scores than PLAN. Notably, even the simplified version of the multi-station phase-picking method, such as the SAGE-based PLAN, outperformed both the single station-based picking methods of EQTransformer and PhaseNet in mF1 scores for S-wave picking. This indicated that the phase-picking accuracy is significantly improved by multi-station picking, which effectively utilizes inter-station contextual information.

We not only adjusted the time threshold while maintaining a constant picking probability for evaluation but also fixed the time threshold (True positive picks were defined as those within 0.5 s of the predicted pick). By changing the probability of picking threshold, we calculated and plotted the precision-recall curves for four models (Supplementary Fig. 2). Given that the curves of Trans-based PLAN consistently appear closer to the upper-right corner, aligning with the results discussed earlier, it is evident that the PLAN model exhibits superior performance in terms of F1 score, encompassing both P-wave and S-wave picking.

Application to seismicity in Japan. We retrained all the methods on the Japan training set for the evaluation. Compared to its performance in the Ridgecrest region, PLAN exhibited an even better performance in Japan (Fig. 3b, c). Further, PLAN demonstrated a remarkably better performance than PhaseNet and EQTransformer for both P- and S-wave picks. The offset predicted by PLAN was notably more accurate than that

predicted by Aggregated-GNN, with a narrower residual distribution (Fig. 3d, e). In terms of depth estimation, although PLAN maintained a narrower residual distribution, the center of the distribution was shifted systematically, compared with the Aggregated-GNN method. Table 1 presents the comprehensive quantitative comparison of the results. Although the TransformGCon-based PLAN method did not demonstrate particular superiority in depth estimation, it excelled in offset estimation (Table 1). Further, the GATv2-based PLAN showed the lowest depth error, indicating potential improvement of localization capabilities of the proposed PLAN.

Similar to the Ridgecrest example, we assessed the phase-picking performance of various models applied to the test data from Japan using mPrecision, mRecall, and mF1 metrics (Table 3) and precision-recall curves (Supplementary Fig. 3). The TransformGCon-based PLAN model achieved superior results in terms of mRecall (95.14 for P-waves and 85.09 for S-waves) and mF1 (95.46 for P-waves and 86.72 for S-waves), whereas EQTransformer performed best in terms of mPrecision of P-waves and S-waves. TransformGCon-based PLAN demonstrated high mRecall scores, indicating that a large proportion of the samples containing P/S-waves were correctly detected. However, this was achieved at the expense of a slightly lower mPrecision compared to that of the EQTransformer, with some non-P/S-waves incorrectly classified as P/S-waves. The mF1 score provided a more comprehensive evaluation of the model performance, considering both the reduction in missed detections and the increase in correct detections. In this context, TransformGCon-based PLAN had the highest F1 score,
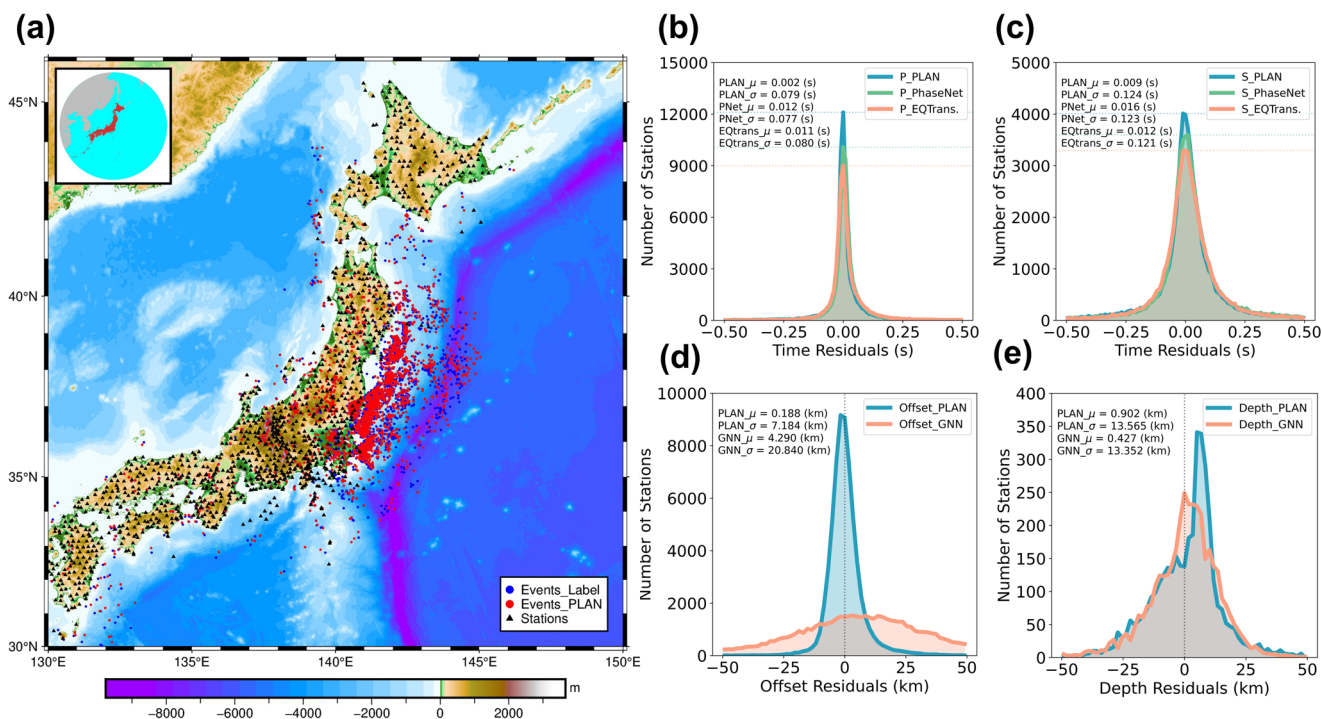
Fig. 3 Distributions of phase picking and location residuals in Japan. a distribution of stations (black triangles) and events of the test data (blue points) used in our study, where the events occurred between 1 January 2011 and 31 December 2011. The red points represent the earthquake locations predicted by PLAN. Similar to Fig. 2**b**, **c** shows the results of P-wave and S-wave arrival time residuals, respectively. **d** and **e** depict the offset and depth residuals between model predictions and the Japan Meteorological Agency (JMA) catalog of located events, respectively.
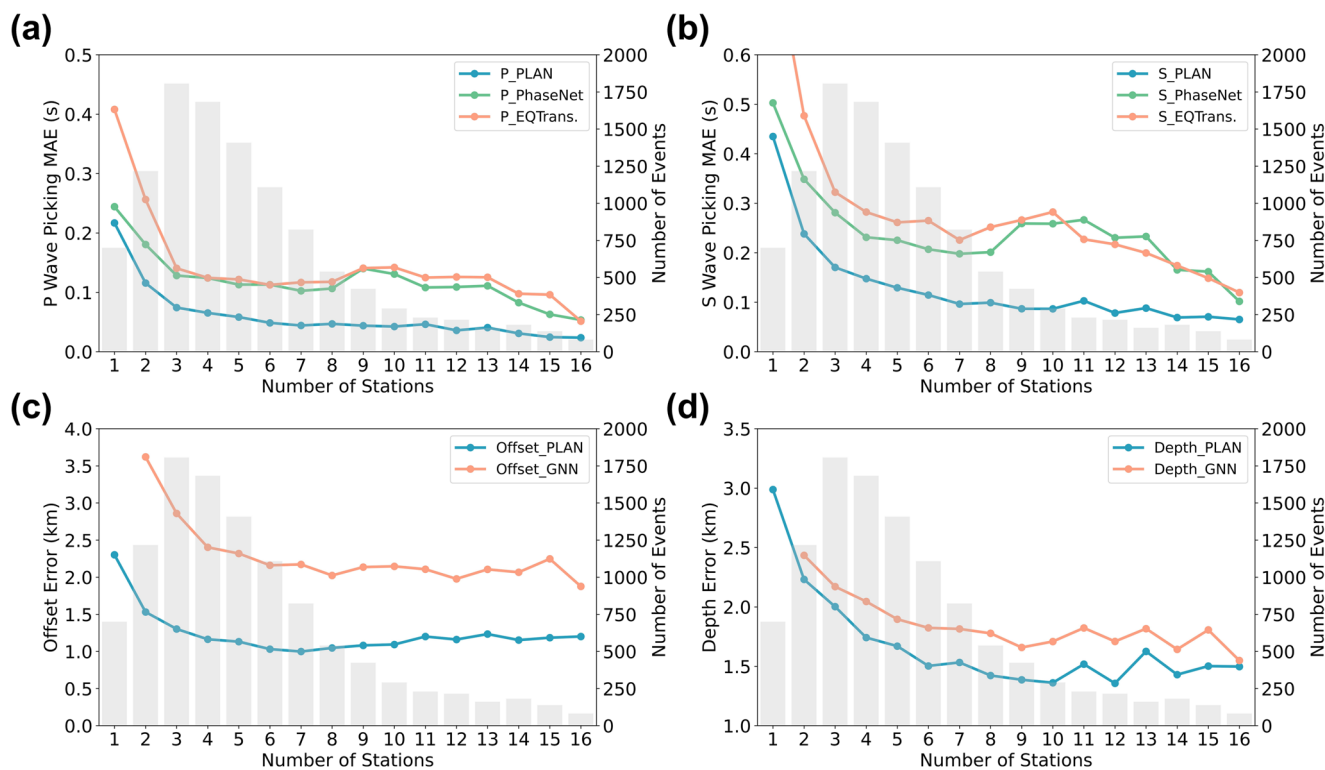

Fig. 4 Comparison of prediction results using different numbers of stations in the Ridgecrest region. The colorful curves in **a**–**d** represent the distributions of prediction errors for P-wave, S-wave, offset, and depth, respectively. The x-axis represents the number of stations, the primary y-axis denotes the prediction errors for phase picking and event localization of stations, and the secondary y-axis represents the number of events recorded by a specific number. Note that for phase picking, prediction residuals of PLAN (blue curves) decrease evidently as the number of stations increases. Moreover, the location errors of PLAN are significantly smaller than those of the Aggregated-GNN method.
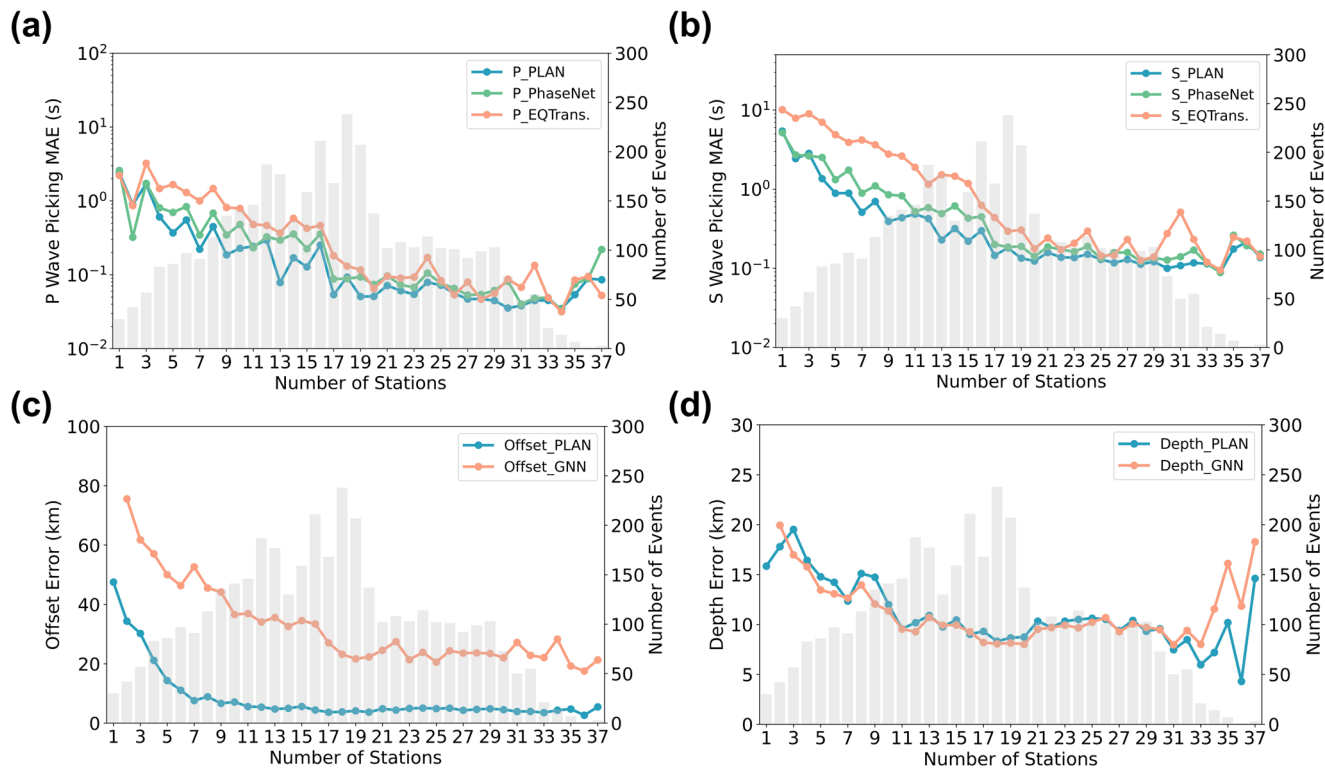
Fig. 5 Comparison of prediction results using different number of stations in Japan. The four distributions (**a**–**d**) are similar to those described in Fig. 4 and the only difference is that we have used logarithmic coordinates for the primary y-axis in **a** and **b**.

Table 1 Location performance in ridgecrest and Japan regions.

Region	Method		Offset MAE (km)		Depth MAE (km)	
			Mean	Std	Mean	Std
Ridgecrest	Aggregated-GNN[36]		2.30	2.30	1.98	1.55
	PLAN	GCN	8.96	6.94	1.43	1.34
		GATv2	8.95	6.90	1.42	1.33
		SAGE	1.28	2.09	1.68	1.42
		Trans	1.09	1.41	1.43	1.33
Hinet	Aggregated-GNN[36]		27.92	26.78	10.30	8.95
	PLAN	GCN	21.30	16.47	13.50	9.95
		GATv2	10.79	10.82	5.59	6.86
		SAGE	4.87	6.11	12.85	9.41
		Trans	4.81	5.83	10.62	8.89

indicating that it effectively reduced the missed detections of P/S-waves and increased the proportion of correct detections.

Application to the continuous waveforms of 2019 Ridgecrest sequence. One key factor in assessing the effectiveness of an earthquake monitoring approach is its ability to process continuous waveform data. PLAN assumes there always exist earthquake signals on every station, and it ignores the case that some stations only have pure noise. As a result, PLAN will issue an earthquake alert even if no earthquake occurs. To address this limitation, we have designed a PLAN-based workflow (Supplementary Fig. 4) that incorporates specific threshold selection procedures, enabling it to effectively handle continuous waveform data and generate an earthquake catalog. (More details of this workflow can be found in the Methods section.)

To assess the performance of this workflow in generating earthquake catalogs, we chose the Ridgecrest earthquake sequence for benchmarking. This choice was motivated by the availability of several well-established earthquake catalogs for this region, making it suitable for comparative analysis. To avoid data leakage, we took the precaution of re-segmenting our training dataset and made some data augmentation (More details of data augmentation can be found in the Methods section). Subsequently, we applied the PLAN-based workflow with the retrained model to process data recorded from July fourth (17:30:00) to July ninth (00:00:00), thereby generating our own earthquake catalog for benchmark comparisons.

We acknowledge that PLAN currently has limitations when dealing with multiple events within a specific time window. In such cases, PLAN focuses exclusively on the first event and outputs the earthquake time and event location corresponding to that event. Consequently, when working with a larger sliding window, there is a risk of missing some events. To address this problem, we implemented a more compact sliding window approach in the continuous waveform data, with each window spanning 30.72 seconds and a 25.72 second overlap between consecutive windows. It's worth noting that this overlap duration is substantially shorter than what is typically employed in other deep learning phase-picking methods designed for single stations. The choice of such a large overlap maximizes the detectability for the dense earthquakes within a short time period. In addition, an example of event processing spanning adjacent time windows can be found in Supplementary Fig. 5.

Since multiple well-established catalogs exist for the 2019 Ridgecrest earthquake sequence, we compare our catalog with them, including SCSN, several deep learning-based methods (such as Liu et al.[11]'s catalog, GaMMA[23] and EQNet[3]), and traditional template matching methods (such as Shelly[49]'s catalog and Ross et al.[50]'s catalog). As there are no ground-truth catalogs

Table 2 Detection Performance in Ridgecrest region.

Method	P picking metrics				S picking metrics		
	mPrecision	mRecall	mF1		mPrecision	mRecall	mF1
PhaseNet[6]	94.83	92.78	93.79		84.50	80.65	82.53
EQtransformer[12]	95.43	91.17	93.25		86.77	78.21	82.27
PLAN-GATv2	95.05	93.02	94.03		85.55	80.49	82.95
PLAN-SAGE	94.99	93.07	94.02		85.65	81.48	83.51
PLAN-Trans	94.65	94.90	94.77		86.88	84.94	85.90

The probability threshold for phase picking is 0.3.

Table 3 Detection performance in Japan.

Method	P picking metrics				S picking metrics		
	mPrecision	mRecall	mF1		mPrecision	mRecall	mF1
PhaseNet[6]	94.87	94.97	94.92		88.26	84.38	86.28
EQtransformer[12]	95.91	94.79	95.35		89.08	84.40	86.68
PLAN-GATv2	95.14	93.81	94.47		88.35	81.87	84.98
PLAN-SAGE	95.65	94.90	95.27		88.45	84.19	86.27
PLAN-Trans	95.79	95.14	95.46		88.41	85.09	86.72

The probability threshold for phase picking is 0.3.

available for continuous waveform data, we employed the same consistent cross-validation test (proposed by Zhu et al.[3]) across different catalogs and utilized the same threshold (3 seconds) for true positive counting. During the cross-validation test, the evaluation metrics are significantly influenced by the number of events. To ensure a fair comparison, we adjusted the threshold, allowing PLAN to generate a catalog with a similar number of events to some of the existing catalogs (a total of 24,270 events). The temporal and spatial distribution of these events can be found in Supplementary Fig. 6.

Supplementary Table 1 presents the precision, recall, and F1-score results of the cross-validation test. These metrics are closely tied to the number of events in the catalog. For example, when considering the SCSN catalog as the benchmark, although Shelly's and Liu et al.'s catalog exhibit higher precision and F1 scores compared to the others, it comes at the expense of lower recall. These catalogs pick fewer events, leaving many true events undetected. In such cases, the higher F1 score may not be a meaningful reference. Additionally, the situation with GaMMA's catalog is similar. While its recall performance is comparable to the catalogs of Ross et al. EQNet, and PLAN, GaMMA's notably high precision is also influenced by the number of events in its catalog. The significant difference in the number of events in GaMMA's catalog, compared to existing catalogs, presents challenges in its evaluation. Therefore, in the cross-validation comparison, we primarily focus on catalogs with a higher number of events.

When considering the SCSN or shelly's catalog as the benchmark, both the PLAN catalog and EQNet catalog exhibit almost identical precision, recall, and F1-scores, all of which slightly surpass the scores associated with Ross et al.'s catalog. Furthermore, when employing Liu et al.'s catalog as the benchmark, the PLAN catalog achieves the highest scores across all three evaluation metrics when compared to other catalogs with similar event quantities. Moreover, when Ross et al.'s catalog and EQNet catalog are employed individually as benchmark, the PLAN catalog consistently outperforms the other catalog in terms of precision, recall, and F1-scores. This indicates that the events detected by the PLAN catalog are consistently present in both Ross et al.'s and EQNet catalogs, implying a high level of credibility and a high likelihood of representing real earthquake events. In conclusion, when processing continuous waveform data, PLAN demonstrates performance that is either on par with or slightly superior to other state-of-the-art methods, consistently yielding a robust and high-quality catalog.

Discussion

PLAN is scalable for accommodating various numbers of stations per event. As PLAN is a network level picking and location model, we further investigated the effect of different numbers of stations on the network performance for phase picking and earthquake location with a test set of the Ridgecrest region (Fig. 4). We calculated the P- and S-wave picking residuals of the three different methods relative to the manual picking results, respectively (Fig. 4a, b). The residuals of the single-station-based picking methods, PhaseNet and EQTransformer, exhibited oscillations for samples with station numbers 3-13 as the number of stations increased. Contrastingly, the residuals of our simultaneous multi-station picking method, PLAN, exhibited a significant residual decrease as the number of stations increased. Although the prediction residuals of the single-station-based methods should not be significantly associated with the number of stations, their prediction residuals still decreased when the number of stations was 13-16. This was probably because the events recorded by more stations tended to be larger and easier to pick.

A comparison of the distribution of prediction errors for earthquake offsets and depths with respect to the number of stations indicated that the errors in PLAN were significantly smaller than those in the Aggregated-GNN method (Fig. 4c, d). However, the errors in offset prediction did not exhibit a significant decrease with an increase in the number of stations. This was likely because a large number of stations would include more distant ones that tended to have large offset prediction errors. As the offset error metric is defined as the average value acquired from multiple stations, an increase in the number of stations can lead to a slightly higher average error for a single event.

Furthermore, the statistical results for Japan (Fig. 5) were similar to those for the Ridgecrest region, with the PLAN method exhibiting smaller phase-picking errors than EQTransformer and PhaseNet, especially for S-wave picking. In addition, as the number of stations increased, the offset prediction error of PLAN became significantly smaller than that of the Aggregated-GNN.

A similar pattern emerges when we compare the distribution of prediction errors in relation to earthquake magnitude and signal-to-noise ratio (SNR) (Supplementary Fig. 7 and Fig. 10). Larger magnitude earthquakes, which are typically detected by more seismic stations, exhibit reduced P- and S-wave picking errors, as well as prediction errors for earthquake offsets and depths, in both the Ridgecrest region and Japan. Additionally, larger SNR, which is usually associated with higher-magnitude events or seismic stations closer to the earthquake's epicenter, also demonstrates smaller picking errors. PLAN, EQTransformer, and PhaseNet show nearly identical performance in terms of P-wave picking errors as earthquake magnitude or SNR increases. Nevertheless, our multi-station picking method exhibits a slight advantage in S-wave picking compared to the other two single-station picking methods, as evident from the narrower distribution of errors (especially in Supplementary Fig. 7b and Figs. 9b and 10b). Moreover, PLAN excels in terms of its location performance, particularly in offset prediction, surpassing Aggregated-GNN (Supplementary Figs. 7c–10c). As earthquake magnitude increases, the offset error for both PLAN and Aggregated-GNN gradually decreases. Nonetheless, at any given magnitude level in both the Ridgecrest region and Japan, the offset error of PLAN remains lower than that of Aggregated-GNN (as depicted in the trends shown in Figs. 4c and 5c).

The ability of our network to handle varying numbers of stations can be attributed to the multi-station association module, which can be separated from the entire network and utilized in a manner similar to the Taup algorithm for estimating the arrival time of earthquakes at stations. Differing from the Taup algorithm, our association module does not depend on an input velocity model. Instead, it empowers the network to comprehend the concept of velocity, enabling the conversion of offsets into relative time shifts. Additionally, unlike the sequential processing of one station at a time in the Taup algorithm, our module simultaneously calculates the time shifts for multiple stations associated with a single event. In essence, our association module can be considered a computationally efficient 3D Taup algorithm that operates without requiring a velocity model. To evaluate the estimation accuracy of the arrival time using this module, we applied the estimated time shifts to align different stations (Supplementary Fig. 11). Because the multi-station association module can accurately estimate the arrival time, the original waveforms from all stations were aligned accordingly.

To further evaluate the estimation accuracy of the arrival time using this module, we employed the TauP algorithm based on the PREM model[51] for comparison (Fig. 6). We also calculated the correlation coefficient (R) between the output of each method and the manually picked P/S-wave time differences. During the training process, this module used the offsets and depths obtained from the earthquake location module as inputs. Therefore, inputting the predicted offsets and depths into this module (Fig. 6d) could yield better P/S-wave time differences than inputting the labeled offsets and depths (Fig. 6c). The multi-station association module with manually labeled offsets and depths yielded less consistent results than the TauP algorithm. This discrepancy may not be solely attributable to errors in the deep learning estimation. Label inaccuracies may have also contributed to this outcome. This assertion was supported by the observation that using the neural network output as the input for the TauP algorithm resulted in greater correlation coefficients

than when label was employed (Fig. 6a, b). Among all the evaluated methods, the estimation results in Fig. 6d show the highest correlation coefficients. Generally, the multi-station association module and the TauP algorithm based on the PREM model have the same level of accuracy in calculating P/S-wave time differences.

Since we have demonstrated that our association module can accurately generate P/S-wave time differences, it is important to note that when processing continuous waveform data, the Taup algorithm cannot reliably estimate wave travel times due to the lack of precise event location information. In contrast, our association module can provide more accurate estimates of P/S-wave time differences, enabling the determination of which stations are associated with the event (Supplementary Fig. 5).

Certain limitations persist within our method, particularly in processing continuous waveform data. While PLAN demonstrates strong performance on test data, it still struggles to provide highly accurate results in offset and depth estimation for continuous waveform data, and it may not match the precision of traditional localization methods. To enhance the accuracy of event localization, incorporating joint relocation algorithms, such as hypoDD[52], into the workflow's final steps would be beneficial. Furthermore, while we employ cross-validation to compare different catalogs in the Ridgecrest region, comparing catalogs that vary in the number of events remains a challenge. Therefore, developing an approach to compare the performance of various catalogs in the absence of ground truth data may be an urgent task for future research.

Moreover, the issue of generalization remains a concern. As the multi-station association module learns velocity concepts during training, pre-trained models from other regions may yield sub-optimal results, particularly in regions with significant velocity variations. Although limited generalization is feasible in smaller geographic areas, such as employing a model trained on Southern California data for earthquake monitoring in Northern California. To address this concern, training region-specific models is a viable approach, as exemplified in our paper, where two distinct models were trained for the Ridgecrest region and Japan.

However, despite PLAN's challenges with generalization, its multi-station association module, which learns velocity concepts during training, offers deeper insights into specific regions. It's important to note that our current association module primarily captures relative velocities. Nevertheless, with adjustments to the labels of shift vectors, it can potentially learn absolute velocities. By inputting hypothetical offsets and depths for each grid in a region, the module could generate corresponding P and S-wave arrival times. Utilizing the time differences and distances between these grids, we can construct a 3D velocity model for a specific region. Thus, while our method may not yet generalize to a broader range of areas, it introduces a concept and possibility in the future: achieving a region's velocity model concurrently with training an earthquake monitoring model using region-specific datasets.

In summary, we present a novel all-in-one multi-task multi-station system called PLAN for earthquake monitoring, which is capable of simultaneous phase picking, phase association, and earthquake location. Unlike current CNN-based methods that perform phase picking station-by-station, phase association, and location separately, our proposed GNN-based multi-station multi-task system best utilizes the inherent inter-task and inter-station constraints. The multi-station association module estimates the phase shift and improves the robustness and accuracy of the phase association process. Eventually, the resulting offsets and depth enables accurate event localization. Our method demonstrates the need to factor mutual constraints among tasks and stations into next-generation earthquake monitoring systems.
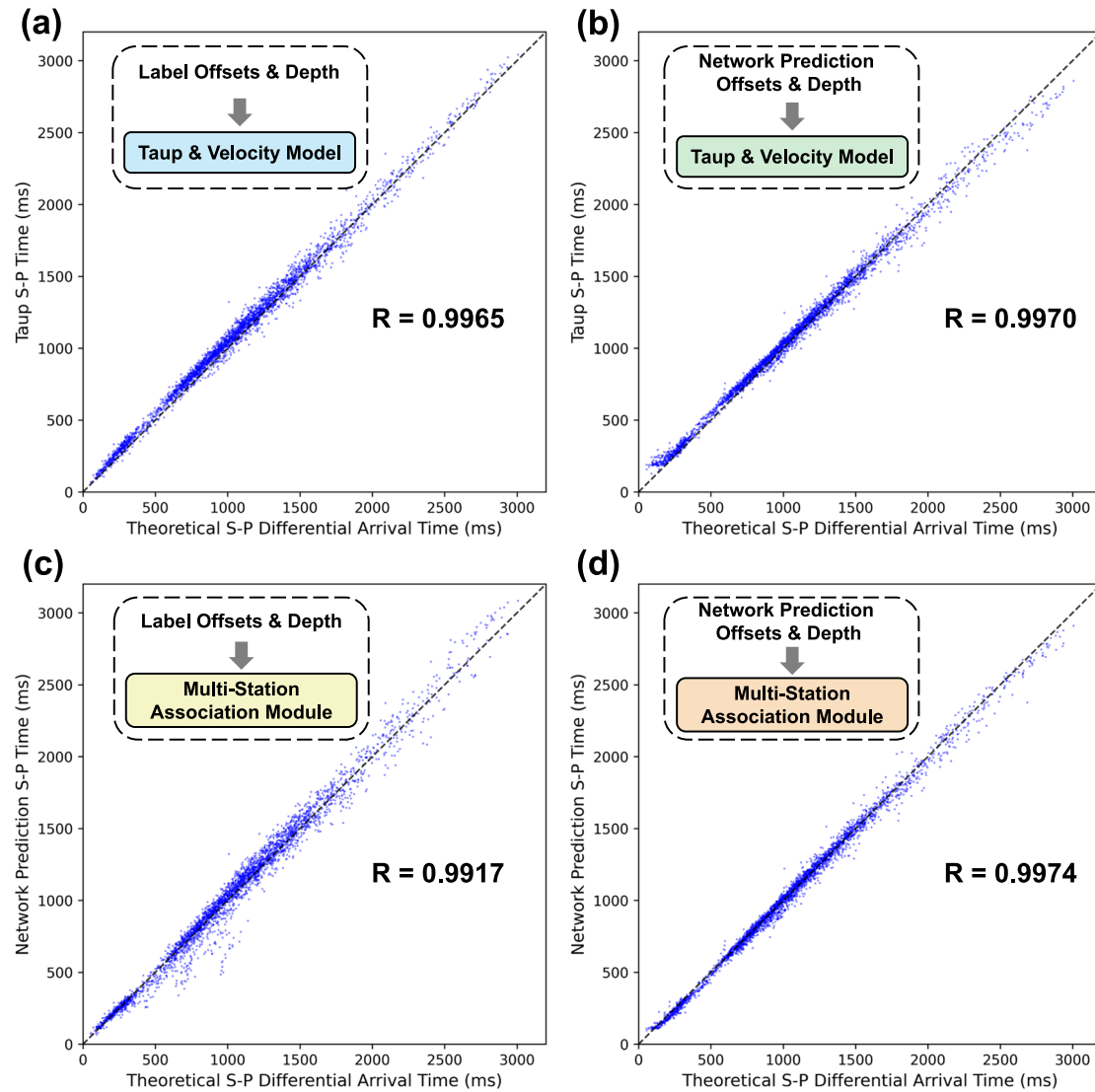
Fig. 6 Comparison of the P/S-wave arrival time estimation using different methods. a and **b** show the crossplots of the S-P differential arrival times computed by the TauP model with the input of offsets and depths from manual labels and network predications, respectively. The *x*-axis represents the manually picked S-P differential arrival times. The *y*-axis represents the S-P differential arrival times obtained by different methods. The crossplots in **c** and **d** show the S-P differential arrival times predicted by our multi-station association module, and they are consistent with those predicted by the TauP model. This indicates that the multi-station association module, detached from the entire trained PLAN model, works physically reasonable compared to the commonly used TauP model.

Methods

Graph based neural network. Several studies have shown that GNNs have the potential to deal with irregularly spaced stations for phase association and event localization[14,16,35–39]. Here, we build a graph-based network (Fig. 1) for mult-station earthquake monitoring. To utilize the GNN, we first need to change the data from the matrix format to the graph format and employ a graph-based representation of the stations, where each station is represented as a node in the graph and the three-channel data and the station location are used as the features of each node. In contrast to the current single-station processing methods[4–12,15], which treat each three-channel data as an individual input sample, our approach inputs all the three-channel data received from multiple stations per event as a single sample. This allows for efficient aggregation of information from multiple stations during network training. As a result, the features of different stations could be effectively integrated using GNNs during the aggregation process.

In this study, we have evaluated various graph aggregation methods, including GCN[40], GraphSAGE[47], GAT[53], GATv2[48], and TransformGCONV[41]. Through this evaluation, we have determined that TransformGCONV, which is based on attention mechanism[42], is the most suitable module for the proposed PLAN. The message aggregation of TransformGCONV could be represented as:

$$\mathbf{x}_i' = \mathbf{W}_1\mathbf{x}_i + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}\mathbf{W}_2\mathbf{x}_j, \tag{1}$$

where $\mathbf{x}_i'$ represents the aggregated features at the source node, and $\mathbf{x}_i$ and $\mathbf{x}_j$ represent the features of the source and distant nodes before aggregation, respectively. $\mathbf{W}_1$ and $\mathbf{W}_2$ are the trainable matrices. In addition, the attention coefficients $\alpha_{i,j}$ are computed via dot-product attention as follows:

$$\alpha_{i,j} = \mathrm{softmax}\left(\frac{(\mathbf{W}_3\mathbf{x}_i)^\top(\mathbf{W}_4\mathbf{x}_j)}{\sqrt{d}}\right), \tag{2}$$

where $\mathbf{W}_3$ and $\mathbf{W}_4$ are the trainable matrices. Similar to the attention mechanism[42], the source feature $\mathbf{x}_i$ and distant feature $\mathbf{x}_j$ are transformed into query vector and key vector, respectively, using $\mathbf{W}_3$ and $\mathbf{W}_4$. Compared to other graph aggregation methods, the use of the attention-based mechanism (equation 1) in TransformGCONV allows for a more fine-grained representation of the relationship between different stations, thereby improving the accuracy and efficiency of the proposed method.

Network Architecture. Here, we design a multi-station multi-task network for simultaneous phase picking, association, and location. The network (Supplementary Fig. 1) comprises four components: a waveform feature extraction module, an earthquake location module, a multi-station association module, and a physics-informed multi-station phase-picking module. Similar to previous deep-learning-based phase-picking approaches[3,6,12,15], we design an encoder to extract waveform features and a decoder to produce phase-picking results. However, to address the multi-station phase-picking problem, we introduce the GNN-based TransformGCONV for aggregating features from multiple stations.

Because aligned waveform features are easily used and aggregated in GNNs for multi-station phase picking, we do not employ it in the waveform feature extraction module (Supplementary Fig. 1a), where the features are relatively shifted in time. Although we use a U-shape neural network for feature extraction to solve the phase-picking problem, it could be replaced with other single-station-based phase-picking networks, such as EQTransformer. No matter what type of network architecture is used, the features extracted from the middle of the network are input into the earthquake location module, and the structure of the final few layers of the network are modified for the purpose of multi-station phase picking. Additionally, the kernel size of all convolutional layers in the waveform feature extraction network is set to 7.

For the earthquake location module (Supplementary Fig. 1b), we first extract features from the normalized coordinate information of the stations within the range of [0,1] through two fully connected layers (3-48-96). Simultaneously, the waveform features extracted from each station are further processed through several convolutional layers and then flattened. Subsequently, the position and waveform features are concatenated and passed through two fully connected layers (192-192-96). This fuses the position and waveform features at each station. The fused features are further aggregated among multiple stations by several GNN layers to predict the offsets of each station with respect to the event and its depth. Because there is only one depth parameter for each sample, we add a global average pooling before the output. In summary, this module allows the integration of both location and waveform information into the feature extraction process, which is crucial for accurate event localization.

Finally, in the physics-informed multi-station phase-picking module (Supplementary Fig. 1d), we incorporate physics-motivated constraints of time alignment among waveforms corresponding to the same earthquake event. We first utilize a mulit-station association module (Supplementary Fig. 1c) to calculate the relative alignment shifts between stations using the estimated offsets and the depth of the event. We then use the shifts to align the waveforms to a common time standard and aggregate the features across multiple stations in the phase-picking module. Subsequently, the aggregated features at each station are unaligned and fed to two layers of convolution to yield final P/S-wave picking results. This process leverages the physical information of the event location to improve the robustness and accuracy of the multi-station phase picking.

Multi-station association module. To simultaneously pick P/S-waves from multiple stations, PLAN utilizes GNNs, which typically aggregates raw signals received at different stations. Feature aggregation across multiple stations introduces inter-station constraints and enhances the features at each station. However, because of different travel times of the same source across different stations, directly aggregating the signals from multiple stations would deteriorate multi-station picking. To address this issue, the proposed method employs a multi-station association module to estimate the time shifts as illustrated in Fig. 1. The input to this module is the offset of each station with respect to the event and its depth. The module output is the corrected time shifts of the P/S-wave for each station. Using the criteria, the multi-station association module was trained to estimate the corrections of P/S-wave for each station. These corrections are then used to align the waveform features, enabling the graph convolution to aggregate the features in a temporally aligned space. Consequently, the method could enhance or compensate for the features at each station by fusing the aligned features from other stations, allowing simultaneous and accurate multi-station picking.

To assess the impact of the Multi-station association module on our picking results, we conducted a comparative analysis by training a neural network without this module and contrasting its performance with PLAN (Supplementary Fig. 12). Across all sub-figures, it is consistently evident that the method without the association module exhibits more pronounced overfitting, resulting in higher loss and lower accuracy compared to PLAN. These findings affirm the efficacy of the introduced multi-station association module in enhancing the network's picking performance and bolstering the model's overall robustness.

The multi-station association module can be utilized independently after training. It converts the distance and depth information into arrival information and calculates the S-P differential travel time[54]. Supplementary Fig. 13 illustrates the arrival time differences of the P/S-waves at various stations in Japan. Although the training process utilizes a maximum of 37 stations for a single event, the module can be adapted to cases with any number of stations (e.g., hundreds of stations shown in Supplementary Fig. 13a–d) to estimate the P/S-wave arrival time differences for all stations. These results indicated that the module effectively enforced physical constraints based on time shifts within the overall network.

Workflow of processing continuous waveform. To enable the application of PLAN to continuous waveform data, we have designed a PLAN-based workflow that incorporates specific threshold selection procedures, allowing for the subsequent generation of a catalog (Supplementary Fig. 4). To ensure the stability of PLAN and its applicability to processing the continuous waveform of the 2019 Ridgecrest sequence, we conducted a retraining process by re-segmenting our training dataset. This process involved excluding data recorded during earthquakes that occurred between July 4th, 2019, and July 9th, 2019. Additionally, during the training process, we introduced data augmentation techniques, including random shifts applied collectively to the waveform windows of all stations. The maximum allowable shift was set to 2 seconds. This approach aimed to improve the model's adaptability to variations in picking positions within the waveform windows extracted from continuous data. Once the retraining was completed, we were able to utilize the retrained

model for processing continuous waveform data using the following steps:

(1) Initial prediction: We apply the retrained PLAN model to the continuous waveforms to obtain initial phase picks in the overlapped windows.

(2) Shift and stack: We use the responses of P and S wave on multiple stations (i.e., the output of the Physics-Information Multi-Station Phase Picking Module in Fig. 1) to detect potential events. Similar to the "shift and stack" strategy widely used in array seismology, we shift the responses to the origin time using the P and S wave shifts (i.e., predicted by the Multi-Station Association Module) and stack them. If the stacked P/S wave response exceeds a threshold (2.4 and 1.2 for P and S wave), the window is expected to contain an event occurring when the stacked response peaks.

(3) Station selection: For the events detected by "Shift and stack", we need to use the picks on different stations to further locate the events. The stations need to be filtered as not every station has picks in the window; not every pick in the window is associated with the origin time estimated by Shift and stack. Only those picks near the theoretical moveout can be used to locate the event. Specifically, the shifted P and S pick should simultaneously exceed a probability threshold (0.24 and 0.12 for P and S wave) and be near the origin time (< 2 s). If > 4 stations meet the criteria, the detected event can be further located.

(4) Catalog generation: In the final step of our process, we input all the candidate stations into PLAN to predict the earthquake time and hypocenter. The earthquake time is calculated using the commonly assumed uniform velocity model of 6 km/s for P waves and 3.4 km/s for S waves in the Ridgecrest region. Simultaneously, the hypocenter is estimated through triangulation using the predicted offsets and depth (i.e., predicted by the Earthquake Location Module). After processing all the overlapped windows, we get a preliminary earthquake catalog from the continuous waveforms. Due to the large overlap, events can be detected by PLAN more than once. We remove the duplicate detections by limiting the minimal separation between consecutive occurrences to 2 s.

Loss function and training details. Our multi-task learning network model has three output results corresponding to phase picking, phase association, and earthquake localization. To train the model, we define three different loss functions for these three different tasks. For phase picking, instead of using the commonly used cross-entropy, we choose the mean square error (MSE) as the loss function, which is suitable for training in multi-task problems. To estimate offsets and depth, which is similar to event localization, we also use MSE as the loss function, as suggested by previous studies[36,39]. Finally, to calculate the P/S-wave shift, we define the loss function as follows:

$$\mathcal{L}_{\Delta p} = \sum_{i=1}^{n} |\mathrm{CTime}_p - \left(\mathrm{label}_{p_i} + \Delta t_{p_i}\right)|, \quad (3)$$

$$\mathcal{L}_{\Delta s} = \sum_{i=1}^{n} |\mathrm{CTime}_s - \left(\mathrm{label}_{s_i} + \Delta t_{s_i}\right)|, \quad (4)$$

where CTime_p and CTime_s represents the reference times where the P- and S-wave picks are aligned, respectively. The reference times for the P/S-wave features were set at the 10 and 15 s in the Ridgecrest region and at the 20 and 32 s in Japan. Additionally, to process continuous waveform data, the reference time for both

the P/S-wave features in the retrained model for the 2019 Ridgecrest sequence processing was set to 0 seconds. Moreover, label_{p_i} or label_{s_i} represents the manually picked P/S-wave arrival time for each station, and Δt represents the predicted P/S-wave shift. Finally, we combine the three types of loss functions to form the overall objective function:

$$L_{\mathrm{total}} = \lambda_1 \mathcal{L}_{\mathrm{picking-p}} + \lambda_1 \mathcal{L}_{\mathrm{picking-s}} + \lambda_2 \mathcal{L}_{\Delta p} \\ + \lambda_2 \mathcal{L}_{\Delta s} + \lambda_3 \mathcal{L}_{\mathrm{offset}} + \lambda_3 \mathcal{L}_{\mathrm{depth}}. \quad (5)$$

Here, we set the coefficients λ_1, λ_2, and λ_3 to 1.

During the training process, the model was optimized using the ADAM[55] method with an initial learning rate of 0.001, which is gradually decreased with a decay rate of 0.9 every 100 epochs. To enhance the training efficiency, we randomly selected 2048 events from the training set for each epoch, rather than using the entire data. The model was trained for a total of 2000 epochs with a batch size of 16, and the training process required approximately 24 h using 1 NVIDIA Tesla A100 GPU.

Evaluation metrics. In previous studies[3,12], true positive phase picks were defined as those within 0.5 s of the predicted pick. The rest were counted as false positives. Nevertheless, owing to potential errors in the labels of the dataset, such statistical results based on a single threshold may not be reliable. Thus, to better evaluate the performance of algorithms, we introduce new metrics, mPrecision, mRecall, and mF1, which are calculated using multiple thresholds, following previous research[56]. The metrics are defined as:

$$\mathrm{mPrecision} \\ = (\mathrm{Precision@11} + \mathrm{Precision@12} + \cdots + \mathrm{Precision1@50})/40, \quad (6)$$

$$\mathrm{mRecall} = (\mathrm{Recall@11} + \mathrm{Recall@12} + \cdots + \mathrm{Recall1@50})/40, \quad (7)$$

$$\mathrm{mF1} = (\mathrm{F1@11} + \mathrm{F1@12} + \cdots + \mathrm{F1@50})/40. \quad (8)$$

where x@11, x@12, $\cdots$, x@50 are Precision, Recall, or F1 metrics when the thresholds are 11, 12, $\cdots$, 50 samples (corresponding to 0.11 s, 0.12 s, $\cdots$, 0.5 s of time), respectively. These metrics, mPrecision, mRecall, and mF1 reward detectors with better picking results and, therefore, can more reasonably or fairly assess the performance of the different methods.

References

1. Beroza, G. C., Segou, M. & Mostafa Mousavi, S. Machine learning and earthquake forecasting–next steps. *Nat. Commun.* **12**, 4761 (2021).
2. Mousavi, S. M. & Beroza, G. C. Deep-learning seismology. *Science* **377**, eabm4470 (2022).
3. Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P. & Beroza, G. C. An end-to-end earthquake detection method for joint phase picking and association using deep learning. *J. Geophys. Res.: Solid Earth* **127**, e2021JB023283 (2022).
4. Ross, Z. E., Meier, M.-A., Hauksson, E. & Heaton, T. H. Generalized seismic phase detection with deep learningshort note. *Bull. Seismol. Soc. Am.* **108**, 2894–2901 (2018).
5. Ross, Z. E., Meier, M.-A. & Hauksson, E. P wave arrival picking and first-motion polarity determination with deep learning. *J. Geophys. Res.: Solid Earth* **123**, 5120–5129 (2018).
6. Zhu, W. & Beroza, G. C. Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* **216**, 261–273 (2018).
7. Mousavi, S. M., Zhu, W., Sheng, Y. & Beroza, G. C. Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.* **9**, 1–14 (2019).
8. Zhu, L. et al. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 mw7. 9 wenchuan earthquake. *Phys. Earth Planet. Inter.* **293**, 106261 (2019).
9. Pardo, E., Garfias, C. & Malpica, N. Seismic phase picking using convolutional networks. *IEEE Trans. Geosci. Remote Sens.* **57**, 7086–7092 (2019).
10. Wang, J., Xiao, Z., Liu, C., Zhao, D. & Yao, Z. Deep learning for picking seismic arrival times. *J. Geophys. Res.: Solid Earth* **124**, 6612–6624 (2019).
11. Liu, M., Zhang, M., Zhu, W., Ellsworth, W. L. & Li, H. Rapid characterization of the july 2019 ridgecrest, california, earthquake sequence from raw seismic data using machine-learning phase picker. *Geophys. Res. Lett.* **47**, e2019GL086189 (2020).
12. 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.* **11**, 1–12 (2020).
13. Yang, S., Hu, J., Zhang, H. & Liu, G. Simultaneous earthquake detection on multiple stations via a convolutional neural network. *Seismol. Res. Lett.* **92**, 246–260 (2021).
14. Yano, K. et al. Graph-partitioning based convolutional neural network for earthquake detection using a seismic array. *J. Geophys. Res.: Solid Earth* **126**, e2020JB020269 (2021).
15. Zhu, J., Li, Z. & Fang, L. Ustc-pickers: a unified set of seismic phase pickers transfer learned for china. *Earthquake Sci.* **36**, 1–11 (2022).
16. Bilal, M. A., Ji, Y., Wang, Y., Akhter, M. P. & Yaqub, M. Early earthquake detection using batch normalization graph convolutional neural network (bngcnn). *Appl. Sci.* **12**, 7548 (2022).
17. Feng, T., Mohanna, S. & Meng, L. Edgephase: a deep learning model for multi-station seismic phase picking. *Geochem. Geophys. Geosyst.* **23**, e2022GC010453 (2022).
18. Münchmeyer, J. et al. Which picker fits my data? a quantitative evaluation of deep learning based seismic pickers. *J. Geophys. Res.: Solid Earth* **127**, e2021JB023499 (2022).
19. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
20. Arora, N. S., Russell, S. & Sudderth, E. Net-visa: network processing vertically integrated seismic analysis. *Bull. Seismol. Soc. Am.* **103**, 709–729 (2013).
21. Gibbons, S. J., Kværna, T., Harris, D. B. & Dodge, D. A. Iterative strategies for aftershock classification in automatic seismic processing pipelines. *Seismol. Res. Lett.* **87**, 919–929 (2016).
22. Zhang, M., Ellsworth, W. L. & Beroza, G. C. Rapid earthquake association and location. *Seismol. Res. Lett.* **90**, 2276–2284 (2019).
23. Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L. & Beroza, G. C. Earthquake phase association using a bayesian gaussian mixture model. *J. Geophys. Res.: Solid Earth* **127**, e2021JB023249 (2022).
24. Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E. & Heaton, T. H. Phaselink: a deep learning approach to seismic phase association. *J. Geophys. Res.: Solid Earth* **124**, 856–869 (2019).
25. McBrearty, I. W., Delorey, A. A. & Johnson, P. A. Pairwise association of seismic arrivals with convolutional neural networks. *Seismol. Res. Lett.* **90**, 503–509 (2019).
26. Yu, Z. & Wang, W. Fastlink: a machine learning and gpu-based fast phase association method and its application to yangbi m s 6.4 aftershock sequences. *Geophys. J. Int.* **230**, 673–683 (2022).
27. Bakun, W. U. & Wentworth, C. Estimating earthquake location and magnitude from seismic intensity data. *Bull. Seismol. Soc. Am.* **87**, 1502–1521 (1997).
28. Zhang, J. et al. Real-time earthquake monitoring using a search engine method. *Nat. Commun.* **5**, 5664 (2014).
29. DeVries, P. M., Viégas, F., Wattenberg, M. & Meade, B. J. Deep learning of aftershock patterns following large earthquakes. *Nature* **560**, 632–634 (2018).
30. Perol, T., Gharbi, M. & Denolle, M. Convolutional neural network for earthquake detection and location. *Sci. Adv.* **4**, e1700578 (2018).
31. Lomax, A., Michelini, A. & Jozinović, D. An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismol. Res. Lett.* **90**, 517–529 (2019).
32. Mousavi, S. M. & Beroza, G. C. Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations. *IEEE Trans. Geosci. Remote Sens.* **58**, 8211–8224 (2020).
33. Zhang, X. et al. Locating induced earthquakes with a network of seismic stations in oklahoma via a deep learning method. *Sci. Rep.* **10**, 1–12 (2020).
34. Münchmeyer, J., Bindi, D., Leser, U. & Tilmann, F. Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophys. J. Int.* **226**, 1086–1104 (2021).
35. McBrearty, I. W., Gomberg, J., Delorey, A. A. & Johnson, P. A. Earthquake arrival association with backprojection and graph theoryearthquake arrival association with backprojection and graph theory. *Bull. Seismol. Soc. Am.* **109**, 2510–2531 (2019).
36. van den Ende, M. P. & Ampuero, J.-P. Automated seismic source characterization using deep graph neural networks. *Geophys. Res. Lett.* **47**, e2020GL088690 (2020).
37. McBrearty, I. W. & Beroza, G. C. Earthquake location and magnitude estimation with graph neural networks. *arXiv* preprint arXiv:2203.05144 (2022).
38. McBrearty, I. W. & Beroza, G. C. Earthquake phase association with graph neural networks. *Bull. Seismol. Soc. Am.* **113**, 524–547 (2023).
39. Zhang, X., Reichard-Flynn, W., Zhang, M., Hirn, M. & Lin, Y. Spatio-temporal graph convolutional networks for earthquake source characterization. *J. Geophys. Res. Solid Earth* **127**, e2022JB024401 (2022).
40. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* preprint arXiv:1609.02907 (2016).
41. Shi, Y. et al. Masked label prediction: unified message passing model for semi-supervised classification. *arXiv* preprint arXiv:2009.03509 (2020).
42. Vaswani, A. et al. Attention is all you need. *Advances In Neural Information Processing Systems*. Vol. 30, 6000–6010 (Curran Associates Inc., Long Beach, California, USA, 2017).
43. Yu, E. & Segall, P. Slip in the 1868 hayward earthquake from the analysis of historical triangulation data. *J. Geophys. Res.: Solid Earth* **101**, 16101–16118 (1996).
44. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vol. 32, 721 (Curran Associates Inc., 2019).
45. Obara, K., Kasahara, K., Hori, S. & Okada, Y. A densely distributed high-sensitivity seismograph network in japan: Hi-net by national research institute for earth science and disasterprevention. *Rev. Sci. Instrum.* **76**, 021301 (2005).
46. Aoi, S. et al. Mowlas: nied observation network for earthquake, tsunami and volcano. *Earth, Planets Space* **72**, 1–31 (2020).
47. Hamilton, W. L., Ying, R. & Leskovec, J. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 1025–1035 (Curran Associates Inc., 2017).
48. Brody, S., Alon, U. & Yahav, E. How attentive are graph attention networks? *arXiv* preprint arXiv:2105.14491 (2021).
49. Shelly, D. R. A high-resolution seismic catalog for the initial 2019 ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity. *Seismol. Res. Lett.* **91**, 1971–1978 (2020).
50. Ross, Z. E. et al. Hierarchical interlocked orthogonal faulting in the 2019 ridgecrest earthquake sequence. *Science* **366**, 346–351 (2019).
51. Dziewonski, A. M. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
52. Waldhauser, F. & Ellsworth, W. L. A double-difference earthquake location algorithm: Method and application to the northern hayward fault, california. *Bull. Seismol. Soc. Am.* **90**, 1353–1368 (2000).
53. Veličković, P. et al. Graph attention networks. *arXiv* preprint arXiv:1710.10903 (2017).
54. Crotwell, H. P., Owens, T. J. & Ritsema, J. et al. The taup toolkit: flexible seismic travel-time and ray-path utilities. *Seismol. Res. Lett.* **70**, 154–160 (1999).
55. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv* preprint arXiv:1412.6980 (2014).
56. Zheng, T. et al. Clrnet: Cross layer refinement network for lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 898–907 (IEEE society, 2022).
57. Uieda, L. et al. Pygmt: A Python Interface For The Generic Mapping Tools (2021).
58. Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

Acknowledgements

Author contributions

X.S. designed the study, prepared the datasets, implemented the codes, and performed the training and tests. X.W. and Z.L. led the project and analyzed the results. S.W. helped implement the codes. S.W. and J.Z. helped prepare the datasets and review the manuscript. X.S., X.W., and Z.L. wrote the manuscript. All the authors contributed to interpretation of the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-023-01188-4.

Correspondence and requests for materials should be addressed to Xinming Wu or Zefeng Li.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Joe Aslin. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.