EdgePhase: A Deep Learning Model for Multi-Station Seismic Phase Picking

Tian Feng, Saeed Mohanna, Lingsen Meng

1Earth, Planetary and Space Sciences, University of California Los Angeles, Los Angeles, CA, USA

Abstract In this study, we build a multi-station phase-picking model named EdgePhase by integrating an Edge Convolutional module with a state-of-the-art single-station phase-picking model, EQTransformer. The Edge Convolutional module, a variant of Graph Neural Network, exchanges information relevant to seismic phases between neighboring stations. In EdgePhase, seismograms are first encoded into latent representations, then converted into enhanced representations by the Edge Convolutional module, and finally decoded into the P- and S-phase probabilities. Compared to the standard EQTransformer, EdgePhase increases the F1 score by 5% on the Southern California Seismic data set. Performance evaluation of EdgePhase shows its strong generalization ability in real-world applications.

Plain Language Summary Identifying seismic phases from continuous waveforms is an important task for earthquake monitoring and early warning systems. Traditional phase recognition methods include visual inspection and detections based on mathematical functions (e.g., STA/LTA, kurtosis, AIC). Recently, machine learning technology has been applied to this task because of its fast operation speed and complete automation. A variety of neural-network-based models take the waveforms of a single station as input and predict the P-phases and S-phases. In this study, we improve the model performance by taking into account the mutually consistent features in multiple stations. We incorporate a Graph Neural Network module to exchange information relevant to seismic phases between neighboring stations. Compared to the standard single station model, our multi-station model performs better on seismic data in Southern California in terms of the precision and recall rate. We also tested our model on the 2020 M7.0 Greece, Samos Earthquake and found that it detected significantly more aftershocks compared to local catalogs in the first month after the mainshock.

1. Introduction

The recent expansion of seismic data and computing resources enables flourishing applications of deep learning (DL) in seismology (Bergen et al., 2019; Bianco et al., 2019; Civilini et al., 2021; Kong, Inbal, et al., 2019; Kong, Trugman, et al., 2019; Kuang et al., 2021; LeCun et al., 2015; Li et al., 2018; Meier et al., 2019; Mousavi & Beroza, 2019, 2020; Nakano et al., 2019; Oord et al., 2016; Perol et al., 2018; Reynen & Audet, 2017; Ross et al., 2019; Saad & Chen, 2020; Saad et al., 2021; Schmidhuber, 2015; Smith et al., 2022; Tian et al., 2020; Valentine & Trampert, 2012; T. Wang et al., 2021). Many of these studies aim to automatically pick P and S arrivals, especially the signals of microseismicity buried under noises. Dozens of DL models prove to be efficient in detecting phases of local events (epicentral distance smaller than 300 km) (Baker et al., 2021; Dokht et al., 2019; Liao et al., 2021; Mousavi et al., 2020; Ross, Meier, & Hauksson, 2018; Ross, Meier, Hauksson, & Heaton, 2018; J. Wang, Sun, et al., 2019; Woollam et al., 2019; Xiao et al., 2021; Y. Zhou et al., 2019; W. Zhu & Beroza, 2019; L. Zhu, Peng, et al., 2019). Some of these DL models exhibit performance close to manual phase picking with precision and recall rates exceeding 95%. Moreover, they are advantageous in terms of their high sensitivity, computation efficiency, and automatic implementation (Mousavi et al., 2020). Compared to conventional automatic phase picking methods, for example, short-term average/long-term average and auto-regression with Akaike Information Criterion (Akazawa, 2004; Allen, 1978; Sleeman & Van Eck, 1999; Stevenson, 1976), detections by DL models show higher accuracy, especially in high-noise-level environments (Mousavi et al., 2020; W. Zhu & Beroza, 2019).
Previous models mainly predict phases with seismograms from a single station. However, some phases are difficult to identify at a single station due to a low signal-to-noise ratio (SNR; Xiao et al., 2021). In the manual phase picking process, humans often rely on waveform consistency between multiple stations to decide if an ambiguous case should be marked as a seismic phase. Therefore, a multi-station approach utilizing the waveform consistency between neighboring stations should improve the performance of DL models in phase picking. Several DL models have attempted to utilize waveforms from multiple stations for source characterization tasks (e.g., earthquake detection, phase-picking, earthquake location). These models typically learn the representations of waveforms in a seismic network through feature extraction or mathematical transformations. The representation learning compresses the high-dimensional raw data to low-dimensional vectors, making it easier to discover patterns and anomalies in the downstream tasks. One approach taken by several studies is to jointly convert seismograms of all stations into latent representations (simplified representation of the input data; Kriegerowski et al., 2019; Yang et al., 2021; X. Zhang, Zhang, et al., 2020; Zheng et al., 2020). This approach is straightforward and easy to implement with conventional DL modules, such as convolutional neural network (CNN; LeCun et al., 1999) and recurrent neural networks (RNN; Hopfield, 1982). However, models using this approach are trained and tested on the same station configuration and region, which limits their generalization capability to other station networks or tectonic settings.

An alternative approach to utilize multi-station data is to combine the feature embeddings from individual stations with an aggregation module (van den Ende & Ampuero, 2020; Xiao et al., 2021; W. Zhu et al., 2021). For example, the Siamese Earthquake Transformer (S-EqT) model aggregates the feature embeddings from station pairs with a cross-correlation module (Xiao et al., 2021). S-EqT has strong generalization ability because the aggregation module is designed to accommodate different station configurations. However, since the communication is limited between station pairs, the waveform consistency of multiple stations is not fully explored.

One solution to establish communications between multiple stations is to utilize the graph neural networks (GNN; Gori et al., 2005; Scarselli et al., 2008; J. Zhou et al., 2020). GNN freely exchanges information between multiple stations, which can handle the irregular structures of graphs with a fixed model architecture (Zhang, Cui, & Zhu, 2020). It has been successfully implemented in seismic source characterization (McBrearty & Beroza, 2022; van den Ende & Ampuero, 2020) and phase association problems (McBrearty & Beroza, 2021). In the multi-station phase-picking task, the high dimensionality of the data makes training a GNN model from scratch difficult. One solution to solve this issue is fine-tuning, a process that adopts a pre-trained model for one given task and then tunes or tweaks the model to adapt a second similar task. It accelerates the training process and overcomes the problem of small data sets. For instance, the aforementioned S-EqT model is fine-tuned on EQTransformer, a popular DL model for the single-station phase-picking task (Mousavi et al., 2020). Here, we attempt to integrate the GNN with EQTransformer to improve the detection ability in low SNR conditions. More specifically, we fine-tune the EQTransformer with a GNN variant named Edge Convolution (Y. Wang, Sun, et al., 2019), and we named this model EdgePhase. In EdgePhase, the seismograms of each station in the seismic network are first encoded into the latent representations individually by the encoder from EQTransformer. Then the edge convolution module combines the latent representations of the target and a neighboring station into a “message.” The “messages” are then used to aggregate the latent representation of the target station into “enhanced representation.” Finally, the “enhanced representation” is decoded into the final P- and S-phase probability functions by decoders from EQTransformer.

The rest of this article is organized as follows: In Section 2, we introduce the EdgePhase model, which is composed of the EQTransformer and an Edge Convolutional module. In Section 3, we describe the Southern California data set used for fine-tuning. In Section 4, we compare the performance of the EdgePhase model with two baseline models. In Section 5, we apply the EdgePhase model in a case study of the 2020 M7 Samos earthquake in Greece. In Section 6, we discuss the seismotectonics of the Samos Earthquake, advantages and potential developments of EdgePhase, and its comparison with other multi-station methods.

2. Model

2.1. EQTransformer

The EdgePhase system is built by integrating the EQTransformer model with the Edge Convolutional module. EQTransformer is one of the most popular DL models for the single-station phase-picking task, and it is also the
benchmark model that other models are usually compared with (Mousavi et al., 2020). This state-of-the-art model picks P- and S-phases with precision close to manual picks by human analysts. It utilizes plenty of deep-learning modules, like 1D convolutions, bi-directional and uni-directional long-short-term memories (LSTM), residual connections, feed-forward layers, transformer, and self-attentive layers. The transformer it adopts has been widely used in the field of natural language processing and computer vision (Devlin et al., 2018; Vaswani et al., 2017). To learn more about these DL modules, please refer to Mousavi et al. (2020). The transformer weights the significance of each part of the input data differentially, and highlights the information relevant to the target. In the phase-picking task, the attention of the transformer is focused on durations, P- and S-phases of events in the seismograms. EQTransformer is built with a multiple-task structure that consists of a deep encoder and three independent decoders (Detection/P-phase/S-phase). In our experiments, we keep the encoder and two decoders (P-phase and S-phase) with their pre-trained weights (328.1 K parameters in total) based on the STanford Earthquake Data set (STEAD) (Mousavi, Sheng, et al., 2019). The encoder encodes 3-component seismograms (X; dimension: 3*6000) into the latent representations (V; dimension: 32*64) (Figure 1a). The Edge Convolutional module transforms the latent representations into the enhanced representations (V’; dimension: 32*64), by transferring the information (messages) related to P- and S-phases between neighboring stations. This is a light module with only 9.3 K trainable parameters, and we will introduce its details in the next section. The two decoders transform latent representations or enhanced representations into P- and S-phase probabilities respectively.

2.2. Edge Convolution

We attempt to improve the phase detection rate in low SNR conditions by incorporating Edge Convolution operations into EQTransformer. Edge Convolution, a variant of graph message passing processing (Gilmer et al., 2017), has achieved great success in classification and segmentation problems (e.g., Y. Wang, Sun, et al., 2019). Edge Convolution enables exchange of information between neighboring nodes, which enhances the relevant signals shared by adjacent nodes. The graph message passing is composed of three steps: (a) the target node gathers messages or feature embeddings from its neighboring nodes; (b) all messages are aggregated via an aggregate function (e.g., sum, max, mean); (c) the pooled messages are then passed through an update function to renew the embeddings of the target node.
In the phase picking task, we represent the recordings of each earthquake or noise sample with a graph $G = (V, E)$, in which $V = \{v_i | i = 1, \ldots, N\} \subseteq \mathbb{R}^{C \times F}$ is the node set with $N$ elements. Each node represents a seismic station, with feature embeddings (channel number is $C$ and embedding length is $F$) exacted from the seismograms. The $E = \{e_{ij}(v_i, v_j) \text{ are connected}\}$ is the edge set containing all edges of $G$. One edge $(e_{ij})$ represents a path used to send messages $(m_{ij})$ between two connecting nodes. The message in this task is a vector that indicates the likelihood that a phase arrives. Based on the assumption that the waveforms of neighboring stations are typically more similar than distant ones, we build an edge between two nodes if their represented stations are considered to be close (geographic distance <100 km). The graph includes a self-loop, meaning each node also points to itself with an edge. An important characteristic of graphs is that they are not defined by the ordering or positioning of the data but only by the relations between data (van den Ende & Ampuero, 2020). The edge convolution operation first collects all the messages associated with all the edges emanating from each node, and then applies a channel-wise symmetric aggregation operation (here we use maximum operation) to update features for each node. By treating each message as a row vector, we can align them all row by row. The maximum operation takes the maximum value for each column, and returns a row vector that matches the message's dimension. Taking node $v_i$ as an example in Figure 1b, we see that it receives four messages $m_{11}$, $m_{12}$, $m_{13}$, and $m_{14}$ from neighbor $v_{11}$, $v_{12}$, $v_{13}$, and itself $v_i$. Then target node $v_i$ will update its value with $\max(m_{11}, m_{12}, m_{13}, m_{14})$. The number of neighbors may be different in other nodes, but a similar operation can be done for every node in the graph.

In order to incorporate Edge Convolutional module into the existing architecture of EqTransformer, the message is designed to be a differentiable function $m_{ij} = h_{ij}(v_i, v_j)$, where $h_{ij}$ is the nonlinear function with a set of learnable parameters $\theta$. We represent $h$ with a neural network composed of five one-dimensional convolutional (conv1D) blocks (Figure 1c). Each block has a convolutional layer (kernel size = 3, stride = 1, padding = 1), a gaussian error linear units activation layer (GELU; Hendrycks & Gimpel, 2016), and a dropout layer (Srivastava et al., 2014) for regularization (dropout rate = 0.1). We concatenate two latent representations along the channel dimension as the input ($\mathbb{R}^{C \times F}$) of the neural network $h_{ij}$. Since the input dimension of the EqTransformer decoder is $C \times F$, we set the output channel number of the 5 conv1D blocks as [2C, 2C, C, C, C], respectively. Compared to the original version of the Edge Convolution module used in the DGCNN model (Y. Wang, Sun, et al., 2019), we made the following changes to adapt to the phase-picking task. In the EdgePhase, a CNN instead of a Multilayer perceptron (MLP) is used to construct messages. This is because the station embedding is a large-size two-dimensional vector, using a CNN instead of a MLP can significantly reduce the number of trainable parameters and skip the flattening operation. Traditional Edge Convolution defines neighbors based on the relative distances with k-nearest neighbors (k-NN). However, only absolute distances can guarantee the coherency of waveforms, because the waveform coherence decays dramatically in an exponential fashion with interstation distance and frequency (Langston, 2014; Luco & Wong, 1986; Zerva & Zervas, 2002). Therefore, unlike k-NN, EdgePhase defines neighbors with the fixed threshold of the geographic distance. The DGCNN model is stacked with many layers of Edge Convolutional layers due to the large number of nodes in graphs. Considering that the number of stations in a network is usually around dozens or hundreds, EdgePhase only uses one Edge Convolutional layer.

3. Data Set

Our training and testing data set (SCSN2021) is built with the continuous waveforms and earthquake catalog of the year 2021 recorded by Southern California Seismic Network (SCSN; Hutton et al., 2010). Here, each sample event (earthquake or noise) is treated as an edgeless graph. Since the number and sequence of stations might vary with different samples, we store the node information of each sample in the dictionary data structure. The node information is not stored in the data set, but constructed during the training, validation, and testing process, because edges could change when performing data augmentation (e.g., reordering, resampling stations; W. Zhu et al., 2020). In addition, the edgeless structure allows for the exploration of different methods and threshold values to establish edges in further research.

In our data set, the 12,718 earthquake samples with more than five labeled phases are selected from 22,619 events that occurred within the year of 2021 in the SCSN catalog. The 15,813 noise samples are randomly selected in the gaps between 5 min before and 40 min after events in the SCSN catalog. Compared to data sets suitable for single-station models (e.g., STEAD with 1.3 million samples), our data set suffers from the problem of a small training data set (28.5k samples), because each sample in our data set is a graph containing all the seismograms.
of one event from a seismic network, rather than one seismogram at a single station. A small set of training samples makes training deep learning models from scratch difficult, so we decided to fine-tune the EQTransformer model with the SCSN2021 data set. The pre-processing steps of seismograms are the same as STEAD, including detrending, bandpass filtering the seismograms to 1–45 Hz, and resampling to 100 Hz.

Previous single-station models achieved robust performance in high SNR conditions, so we intentionally make our training data set noisy to test the performance of multi-station models. STEAD only keeps the seismograms with both P- and S-phase labels, but our data set includes seismograms with a P-phase label or S-phase label. The seismograms with only one phase label tend to be of low SNR as the other phase is not recognized by the SCSN phase picking system. Compared with samples in STEAD, the waveforms of samples in the SCSN2021 are more noisy. The median values of the signal-to-noise ratio of the STEAD test set and SCSN2021 test set are 25.0 and 16.3 dB, respectively (Figure S1 in Supporting Information S1). The SNR is calculated in the same way as in the STEAD data set (Mousavi, Sheng, et al., 2019). For the prediction targets (labels), we convert the phase arrival times into a temporal distribution with a triangular shape, just like EQTransformer. If one station has multiple types of channels (e.g., Broadband, High Gain), we consider them as multiple stations with the same geographical location. If some components or segments of waveforms are missed, we pad the gaps with zeros. Finally, we split the data sets into training, development, and test sets, which account for 80%, 10%, and 10% of the total samples, respectively.

4. Performance

To understand the contribution of the Edge Convolutional module to the overall system of EdgePhase, we compared the EQTransformer model with Edge Convolutional module (EdgePhase) and the standard EQTransformer without Edge Convolution (Baseline-A). For a fair comparison, EdgePhase and Baseline-A are both fine-tuned with the SCSN2021 training set. To study the effect of fine-tuning, we also added Baseline-B to the comparison, which is the standard EQTransformer trained with the STEAD data set without the fine-tuning. Due to the consistency of region and station configuration between training and test set, the EdgePhase and Baseline-A are expected to achieve better performance than Baseline-B on the SCSN2021 test set. In the fine-tuning process of EdgePhase and Baseline-A, we adopt the binary cross-entropy and Adaptive Moment Estimation (Adam) as our loss function and optimizer, respectively. The training loss was checked every 5,000 training steps. The learning rate for the optimizer started from 10e−4 and decreased by 90% every time the training loss stopped decreasing in the past 10 checks. We update the weights of the whole network in a batch size of 16. Each sample may have a different number of stations, so we make the data generator sampling up to 32 stations per training sample. The EdgePhase and Baseline-A models start to converge at 52k and 40k training steps, respectively.

We define the peak in the probability function above a certain detection threshold as a phase pick, with the phase arrival time located at the center of the peak. The pick is counted as a true positive (TP) if the time residual of matched labeled phases is smaller than 0.5 s. If the pick can't match any labeled phases, it is deemed a false positive (FP). If a labeled phase does not match with any picks, then it is considered false negative (FN). Based on TP, FP, and FN values we can calculate three basic performance metrics of machine learning models: precision \( \frac{TP}{TP+FP} \), recall \( \frac{TP}{TP+FN} \), and F1 score \( \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \). For the true positives, we evaluate the accuracy of picking time by calculating the mean, standard deviation (STD), and mean absolute error (MAE) of their residuals with the ground truth (Table S1 in Supporting Information S1).

To determine the model detection thresholds, we use the F1 score, which is the weighted average of precision and recall. We set the thresholds of the P-phase for EdgePhase, Baseline-A, and Baseline-B at 0.39, 0.64, and 0.05, when they achieve maximum F1 scores at 0.88, 0.84, and 0.66, respectively (Figure 2; Table S1 in Supporting Information S1). Similarly, we set the thresholds of the S-phase for EdgePhase, Baseline-A, and Baseline-B at 0.28, 0.4, and 0.03, when they achieve maximum F1 scores in the test set at 0.86, 0.81, and 0.62, respectively. The Baseline-B model achieves the F1 score of 0.99 and 0.98 for P and S-phases on the STEAD test set, which are higher than on the SCSN2021 test set (0.66 and 0.62). This is due to consistency of region and station distribution between training and test set, and the earthquake waveforms in the SCSN2021 test set being more noisy than STEAD, as we laid out in the previous section. In addition, the STD, and MAE on phase picking time residuals by EdgePhase increase by about 0.01 s, compared to two other models. This is also reasonable as the additional detections made by EdgePhase tend to have blurry phase arrivals, which would contribute to larger picking errors.
time residuals. In addition, the geographic distance threshold for constructing edges is discussed in the supplementary materials (Text S1 in Supporting Information S1).

In previous analysis, EdgePhase outperforms two baseline models with higher precision and recall rate. To study the effect of SNR on the phase picking performance, we compare the number of TP, FN, and FP generated by the three models at low (0 db), medium (10 db), and high (20 db) SNR bands (width: ± 1 db). The P-phase TP (matched phases) numbers of EdgePhase, Baseline-A, and Baseline-B models are 183, 109, 5 at low SNR; 1048, 995, 468 at middle SNR; and 988, 959, 915 at high SNR, respectively (Figure S2 in Supporting Information S1). The P-phase FN (missed phases) numbers of EdgePhase, Baseline-A, and Baseline-B models are 443, 517, 621 at low SNR; 113, 166, 693 at middle SNR; and 39, 69, 112 in high SNR, respectively. The P-phase FP (false alarms) numbers of EdgePhase, Baseline-A, and Baseline-B models are 36, 22, 14 at low SNR; 39, 36, 76 in middle SNR; and 17, 22, 155 in high SNR, respectively. Compared with two baseline models, EdgePhase detects more and misses fewer phases at all three SNR bands. EdgePhase also makes fewer false alarms at high SNR conditions, but more false alarms at low SNR conditions. Compared with two baseline models, EdgePhase is more sensitive to phase arrivals at low SNR conditions by making significantly more phase picks, so it is also reasonable to make more false alarms. It demonstrates EdgePhase's superior detection capability at the environment of 0–20 db SNR. Based on Figure S3 in Supporting Information S1, a similar conclusion can also be drawn for S-phases. To sum up, EdgePhase outperforms two baseline models, especially at low to medium noise levels (SNR between 0 and 20 db).

Next, in Figure 3, we show the detection probability functions of the three models with an example of an M 1.43 earthquake in the test data set that occurred on 25 April 2021 (202110425M1). The EdgePhase model outperforms two baseline models by detecting all labeled P- (11) and S-phases (9). The Baseline-A detects 5 P- and 4 S-phases, while Baseline-B detects 4 P- and 7 S-phases with 6 false detections. Some of the missed detections in Baseline-A show a peak in the probability functions, but smaller than the detection threshold. Baseline-B sometimes confuse P-phases with S-phases and there are no corresponding peaks to the missed detections in its probability functions. In general, EdgePhase outperforms Baseline-A, and Baseline-A outperforms Baseline-B. To study the efficiency of the Edge Convolutional module in exchanging information relevant to P- and S-phases, we visualize the input (latent representations) and output (enhanced representations) of this module. We visualize the feature map of the three neighboring stations for Event 202110425M1 (Figure S4 in Supporting Information S1), which are sorted in ascending order based on the mean values of the features. For the enhanced representations, the amplitudes of features are either increased or decreased in the region between the P-phases and S-phases. But there are clear changes of amplitudes at the arrivals of the P-phases and S-phases in the feature maps, compared with the seismograms before and after the seismic phases. This phenomena is especially clear for the station SLB. The latent representations show fluctuations before the P-phases but no clear change of pattern before and after P-phase arrivals in many channels. The boundaries between phases and background are clearer in the enhanced representations than in the latent representations, which demonstrates that the communication between stations effectively suppresses the noise before P-phases and after S-phases in the feature spaces.

To conclude, fine-tuning is essential to EQTransformer when applying to new data sets or applications. The Edge Convolutional module, which can convert a single-station model into a multi-station model, makes additional improvements during the fine-tuning process.
5. Case Study of the M7.0 Samos Earthquake

To evaluate the performance of EdgePhase in a real-world application, we examined the aftershock region of the M7.0 Samos (Néon Karlovasion) earthquake which occurred 14 km northeast of the Greek island of Samos on 30 October 2020. It was the deadliest earthquake in 2020, causing more than 100 deaths during the mainshock and tsunami. This devastating earthquake happened due to the north-south extensional stress during slab rollback, where the African plate subducted beneath the Aegean and Anatolian Microplates (Figure S5 in Supporting Information S1; Kassaras et al., 2020; Meng et al., 2021). Preliminary results indicate that the mainshock occurred on the Offshore North Samos (Kaystrios or Samos Basin) Fault, which has not hosted a large earthquake since the mid-eighteenth century (Foumelis et al., 2021; Papadimitriou et al., 2020). The north-dipping candidate plane from the focal mechanism is more likely to be responsible for the 2020 quake, as suggested by the uplift at the western part of Samos Island (footwall), and the over 10 cm of subsidence at the northermost edge of the central part of the island (Evelpidou et al., 2021; Papadimitriou et al., 2020). However, the distribution of aftershocks in regional catalogs (NKUA, GI-NOA, and AFAD) does not show any preference for the nodal plane (Foumelis et al., 2021; Papadimitriou et al., 2020).

Early aftershocks immediately following the mainshock are usually relevant to the mainshock rupture process (Mendoza & Hartzell, 1988; Wetzler et al., 2018), so we study the distribution of aftershocks in the first month after mainshock. The phase picking model was applied to continuous data from 72 stations around the Aegean...
Sea (Figure S5 in Supporting Information S1). These waveforms are not included in the STEAD and SCEC2020 data sets, so they are suitable for testing the generalization ability of EdgePhase. The pre-processing of the raw waveforms included bandpass filtering (1–45 Hz) and detrending. We then divide the continuous waveforms into 60-s sliding windows with a step of 18 s and feed them into the EdgePhase phase picker. The computational time required to process 24 hr of continuous data from 72 three-component seismometers is around 20 min on an economical GeForce-GTX-1070-8 GB GPU. As a result of using low detection thresholds (0.18 for P-phases and 0.15 for S-phases), we were able to detect 979,218 P-phases and 1,000,341 S-phases with relatively low SNRs. Then, we applied the rapid earthquake association and location method (REAL), least squares location method (VELEST), and double-difference earthquake location algorithm (HypoDD), in that order, to perform phase association and event locations (Kissling et al., 1994; Waldhauser & Ellsworth, 2000; M. Zhang et al., 2019). The magnitude estimation method and parameter setting of earthquake location methods are included in the supplementary materials (Hutton & Boore, 1987; Ozer et al., 2018; Tan, 2013; Text S2 and Text S3 in Supporting Information S1).

In total, we located 1,222 events with VELEST and relocated 687 events with HypoDD. Compared to the local catalog (421 events, Figure S6 in Supporting Information S1) from the NKUA, this study (Figure S7 in Supporting Information S1) increases the detections by 190% with a distribution more conformative to a linear fault trace (Figure 4a). It demonstrates that the detection ability of the DL phase picking model is on par with that of the manually reviewed catalog. We consider a DL detection to match with a known event in the NKUA catalog if they occurred within 3 s and 30 km of each other. Under such matching criteria, we matched approximately 53% of the events listed in the NKUA catalog, with the event locations of this study generally being somewhat north of the NKUA catalog (Figure S8 in Supporting Information S1). Based on matched event pairs, we calibrate the estimated local magnitude of our detected events using linear regression (Figure S9 in Supporting Information S1). We then check the 997 new detections found by this study, and find that earthquake signals are quite clear in most traces, especially at close stations (four examples shown in Figure S10–S13 in Supporting Information S1).

6. Discussion

6.1. Seismotectonics of the Samos Earthquake

Similar to the preliminary report by Papadimitriou et al. (2020), we group aftershocks of the M7 Samos Earthquake into four clusters along the fault strike direction (Figure 4a) with the DBSCAN algorithm (distance <4.3 km; at least 43 samples in a group). Events in Clusters 2 and 3 started to appear in an earthquake swarm from 20 June to 30 June 2009. During the period of the swarm, one $M_w$ 5.1 event and more than 80 $M_L > 1.5$ events occurred around Samos Island (Tan et al., 2014). The clustering of events indicates the concentration of stress in these locations as well as the presence of frictional heterogeneities on the fault plane. Furthermore, the slow unlocking process (precursor) of the Offshore North Samos Fault might have existed several years before the 2020 M7 event similar to those observed prior to several megathrust earthquakes (e.g., Huang & Meng, 2018).

The cross-section along the A-A’ direction shows that the events (Clusters 1 and 2) west of the mainshock are shallower than the events east of the mainshock (Clusters 3 and 4; Figure 4b). Furthermore, the events in the west segments are restricted to a depth of 13–17 km. By examining the cross-section along the B-B’ direction, we find that the south-dipping plane of the mainshock by AFAD (dashed lines in Figure 4c) seems to fit the lower bound of clusters 1 and 4 well. Cluster 2 and 3 events are close to the mainshock hypocenter, with shallower depths than the mainshock (Figure 4d). These events are less consistent with the mainshock fault plane, which suggests that aftershocks may involve the triggering and activation of the secondary splay faults.

The results of this case study demonstrate that our fine-tuned DL phase picking method incorporated with traditional earthquake location methods, could automatically monitor earthquakes. Meanwhile, the performance of the multi-station phase-picking model is proven to be comparable to human efforts.

6.2. Advantages and Further Development of EdgePhase

In early phase-picking studies, the seismograms/spectrograms were treated as images by CNNs (Ross, Meier, & Hauksson, 2018; Ross, Meier, Hauksson, & Heaton, 2018; J. Wang, Sun, et al., 2019; Woollam et al., 2019; W. Zhu & Beroza, 2019; L. Zhu, Peng, et al., 2019), or as audio data by incorporating RNNs (Mousavi et al., 2020;
In this study, the seismograms and geographical locations of multiple stations are treated as a graph through the GNN framework. The communication between stations through the Edge Convolutional module brings the phase-picking model into a new stage, where the phase detection not only depends on seismograms from a single station, but also waveform consistency in a seismic network. Compared with the single-station model (Baseline-A), the precision and recall of the multi-station model (EdgePhase) increases by 5%, and the detection accuracy in low SNR conditions is improved.

Although the EdgePhase model is fine-tuned with data in Southern California (SCSN2021), it showed strong generalization ability in the case study of the 2020 Samos earthquake sequence in Greece. There are two reasons...
for the strong generalization ability. First, the encoders and decoders of EQTransformer are pre-trained with a global data set (STEAD). Second, the information exchange mechanism in EdgePhase only requires that the neighboring stations are within a certain distance range, therefore its application is not restricted to a specific station configuration as is the case of the end-to-end DL models (Kriegerowski et al., 2019; Yang et al., 2021; X. Zhang, Zhang, et al., 2020; Zheng et al., 2020).

In addition to strong generalization ability, the Edge Convolutional module can be easily incorporated into any pre-trained phase-picking models in the encoder-decoder architecture. For example, one potential improvement to PhaseNet (W. Zhu & Beroza, 2019) is to add the Edge Convolutional module right before the first devolutional layer. There is no need to change its U-net architecture and residual connections. We can even pick and choose encoders and decoders from different models, with the Edge Convolutional module connecting them. Reusing the pre-developed encoders, exchange modules, and decoders can accelerate the development of new models through transferring learning or fine-tuning.

There is still potential for improvement of EdgePhase in further work. Possible developments include replacing the Edge Convolutional module with a more general message passing network (Gilmer et al., 2017), trying different functions to build and aggregate messages; fine-tuning with a global data set (large size), and exploring different algorithms such as k-NN to define neighbors of stations. In the current EdgePhase model, two stations are either classified as neighbor or non-neighbors based on their relative distance. This rough binary classification ignores the differences between close and far neighbors, and an alternative method is to add a distance feature to the edges connecting the stations. The distance feature can potentially serve as weighting coefficients for the waveform embedding of two connected nodes, although the adding edge features may require more computing resources, which scales with the square of the number of nodes.

The EdgePhase is very suitable for solving node-level tasks, in which the model predicts properties for each node in a graph. Besides the phase-picking task, EdgePhase can also be applied to the seismograms denoising problem. The phase-picking and seismogram denoising tasks transform the raw seismograms into phase probability functions and denoised seismograms for each station, respectively. Like the phase-picking task, separating noise from signals also benefits from communication between stations. Based on the assumption that a similar pattern of the earthquake signal is shared at neighboring stations, the Edge Convolutional module can be modified to transfer the feature embeddings relevant to the earthquake signals. One can adopt encoders, decoders, and loss functions in the previous single-station seismograms denoising models (Novoselov et al., 2020; W. Zhu, Peng, et al., 2019) and fine-tune them as is done in this study.

### 6.3. Comparison to Other Multi-Station DL Models

In this section, we compare the EdgePhase model with three other multi-station DL models that combine the feature embeddings from individual stations with an aggregation module (van den Ende & Ampuero, 2020; Xiao et al., 2021; W. Zhu et al., 2021). van den Ende and Ampuero (2020) aggregate the feature embeddings of all stations with a node-wise maximum function (we will use EA2020 as the abbreviation for their model). For each element of the latent representation, the node-wise maximum function selects the largest value among embeddings of all stations, therefore the aggregation result is not affected by the number or the order of stations. The EQNet by (W. Zhu et al., 2021) adopted a different approach to aggregate feature embeddings related to P- and S-phases through a shift-and-stack operation similar to the backprojection process. The shift operation is performed on P- and S-phase embeddings, which can be regarded as a preprocessing step. The stacking operation is to node-wise sum up the shifted embeddings, which works as an aggregation function. The S-EqT model (Xiao et al., 2021) retrieves phase picks in target seismograms (low SNR) by referencing confident phase picks in the template seismograms (usually has high SNR). Specifically, it first extracts feature embeddings from seismograms with the encoder of EQTransformer. Then it performs enhancing and cross-correlation operations on the feature embeddings of the template and target seismograms. The cross-correlation operation works as an aggregation module to gather information about similarity between template and target stations.

These models attempt to aggregate station embeddings with the knowledge of graphs and earthquake physics (e.g., phase velocity of the P and S arrivals, waveform similarity), which correspond to GNN modules and physics-based modules in models, respectively. An elementary GNN module transforms nodes, edges, and the global context of graphs, while a physics-based module is modified from some widely used multi-station
geophysical technologies or models (e.g., waveform cross-correlation, backprojection, and focal mechanism inversion). Compared to the GNN modules, physics-based modules benefit from additional physical constraints but are only applicable to some specific tasks. The S-EqT and EQNet models include physics-based modules, while EdgePhase and EA2020 only utilize GNN modules.

In addition to the presence of physics-based modules, these four models show differences in methods to aggregate feature embeddings. The EA2020 and EQNet models aggregate feature embeddings of all nodes with a function that satisfies the law of permutation invariant, such as maximum (EA2020) and summation (EQNet). Permutation invariant means that the values of these functions do not change when exchanging the order of the components/elements of input. The aggregation modules of EQNet and EA2020 treat every station equally and transform the feature embeddings of the whole seismic network into a matrix of a fixed size, which does not change proportionally with the number of stations. This design is suitable for solving graph-level tasks, which predict properties for the entire graph. Here, we take the source characteristic task as an example: the amount of parameters involved in predicting magnitude and location only depends on the source (graph) itself, not the number of stations (nodes).

On the other hand, S-EqT and EdgePhase use the message passing mechanism, which applies the aggregation module N times for a N-station network. In each aggregation operation, there is a non-repeating primary station and several reference stations, and the fixed-size output represents the feature embeddings for the primary station. By scaling the number of aggregation operations with the number of stations, the size of all stations' embeddings is proportional to the number of stations, therefore message passing networks are suitable for solving node-level tasks.

Next, we will compare S-EqT and EdgePhase in more detail, because they are both built with a message passing mechanism and for the same task (multi-station phase-picking). In the S-EqT model, a primary station pairs another referencing station with reliable phase picks as its neighbor. In the EdgePhase model, more neighbors are connected with the primary station based on the geographic distance, which means more data is used to predict phases for each station. The geographic locations of the template and searching stations are not considered in the S-EqT model. However, the distances between stations is still useful information for phase picking, since nearby stations have closer phase arrival times on the seismograms for the same event. In addition, S-EqT constructs a message between the primary and referencing station through the cross-correlation between two station embeddings. In the EdgePhase, messages are learnt by a neural network, which takes the embeddings of sending and receiving stations as input. The neural network, rather than simply calculating the similarity between two embeddings, is more expressive than the cross-correlation function. Therefore, a neural network could extract more relevant information that nodes intend to communicate with each other through optimizing the weights to minimize loss function. In the training process, EdgePhase makes two additional improvements compared with S-EqT. First, the parameters of the EqTransformer model are fixed during the training process of S-EqT, but are trainable during the fine-tuning process of EdgePhase. Making EqTransformer parameters trainable increases the degrees of freedom of the model, but achieves better performance after fine-tuning. Second, S-EqT is optimized with Stochastic gradient descent by setting batch size to one for convenience, while EdgePhase is optimized with mini-batch gradient descent (batch size = 16). Stochastic gradient descent updates parameters once per sample, so it usually takes a longer time to converge to the minimum value of the loss function and involves numerous random and less efficient update steps compared with mini-batch gradient descent.

To conclude, although these models have substantial differences in model structures, they all analyze multi-station waveforms in a GNN framework. As we gain a better understanding of constructing graphic networks for geophysical problems, the next generations of GNN models will have a stronger expression and generalization ability. Finally, we expect more developments about how to efficiently exchange and aggregate information among multiple stations, how to design GNN models for different geophysical tasks, and how to build a data set suitable for these models.

7. Conclusions

By integrating an Edge Convolutional module with EQTransformer, we construct a multi-station phase-picking model named EdgePhase. EdgePhase takes into account the waveform consistency between neighboring stations and improves the phase detection capability at low to medium noise levels (SNR between 0 and 20 db). We compared EdgePhase with the fine-tuned EQTransformer (Baseline-A) and the original EQTransformer
(Baseline-B) using the Southern California data set (SCSN2021). The best F1 scores for EdgePhase, Baseline-A, and Baseline-B are 0.88, 0.84, 0.66 in P-phase detections; 0.86, 0.81, 0.62 in S-phase detections, respectively. It demonstrates that fine-tuning the ETransformer with the Edge Convolutional module makes significant improvements to phase picking tasks. The EdgePhase also shows a strong regional generalization ability, as it was trained with the Southern California data set but performs well in an independent data set of the 2020 Samos earthquake sequence. Compared with the NKUA catalog, EdgePhase detects 190% extra events, with a higher fidelity of event locations. With its comparable performance to human seismologists, EdgePhase can automatically detect and monitor regional seismic activities in real-time. Furthermore, the Edge Convolutional module could potentially be applied to other pre-trained phase picking models (e.g., PhaseNet) and other node-level tasks (e.g., multi-station seismograms denoising).

**Data Availability Statement**

The fault data as well as EdgePhase code can be found at github (zenodo) repository (https://github.com/lovelytt0/EdgePhase; https://zenodo.org/record/6961446; YuteFOz01Vp). Our training data is based on the Southern California Data set (https://scedc.caltech.edu/data/cloud.html). The data of the Greece Earthquakes can be downloaded through Obspy FDSN Client KOERI (http://eida.koeri.boun.edu.tr) and NOA (https://eida.gein.noa.gr). Figures were produced with Matplotlib and Obspy. We are grateful to the Seismological Laboratory of National and Kapodistrian University of Athens (NKUA, http://www.geophysics.geol.uoa.gr/stations/maps/recent.html), Geodynamics Institute of the National Observatory of Athens (GI-NOA, https://www.gein.noa.gr/), and Disaster and Emergency Management of Turkey (AFAD, https://en.afad.gov.tr/) for the earthquake catalog. We implement the machine-learning model with the SeisBench API (Woollam et al., 2021) (https://github.com/seisbench/seisbench).

**References**

Akazawa, T. (2004). A technique for automatic detection of onset time of P-and S-phases in strong motion records. In Proceedings of the 15th world conference on earthquake engineering.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5), 1521–1532. https://doi.org/10.1785/0219780051521

Baker, B., Holt, M. M., Pankow, K. L., Koper, K. D., & Farrell, J. (2021). Monitoring the 2020 magna, Utah, earthquake sequence with nodal seismometers and machine learning. *Seismological Society of America, 92*(2A), 787–801. https://doi.org/10.1785/02202000316

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science, 363*(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., et al. (2019). Machine learning in acoustics: A review. *arXiv e-prints.*

Civilimi, F., Weber, R., Jiang, Z., Phillips, D., & Pan, W. D. (2021). Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning. *Geophysical Journal International, 223*(3), 2120–2134. https://doi.org/10.1093/gji/ggaa083

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dokht, R. M., Kao, H., Visser, R., & Smith, B. (2019). Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters, 90*(2A), 481–490. https://doi.org/10.1785/02201803008

Evelpidou, N., Karkani, A., & Kampolis, I. (2021). Relative sea level changes and Morphotectonic implications triggered by the Samos earthquake of 30th October 2020. *Journal of Marine Science and Engineering, 9*(1), 40. https://doi.org/10.3390/jmse9010040

Fournelis, M., Papazachos, C., Papadimitriou, E., Karakostas, V., Ampanatzis, D., Moschopoulos, G., et al. (2021). On rapid multidisciplinary response aspects for Samos 2020 M7.0 earthquake. *Acta Geophysica, 69*(3), 1–24. https://doi.org/10.1007/s11600-021-00578-6

Giller, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *International conference on machine learning* (pp. 1263–1272).

Gori, M., Monfardini, G., & Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings. 2005 IEEE international joint conference on neural networks* (Vol. 2, pp. 729–734).

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415.

Hofield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America, 79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Huang, H., & Meng, L. (2018). Slow unlocking processes preceding the 2015 MW 8.4 ilapel, Chile, earthquake. *Geophysical Research Letters, 45*(9), 3914–3922. https://doi.org/10.1002/2018GL077060

Hutton, L., & Boore, D. M. (1987). The ml scale in southern California. *Bulletin of the Seismological Society of America, 77*(6), 2074–2094. https://doi.org/10.1785/021987008100062074

Hutton, L., Woessner, J., & Hauksson, E. (2010). Seventy-seven years (1932–2009) of earthquake monitoring in southern California. *Bulletin of the Seismological Society of America, 100*(2), 423–446. https://doi.org/10.1785/0120090130

Kassar, I., Kapatridis, V., Ganas, A., Tzanis, A., Kosma, C., Karakonstantis, A., et al. (2020). The new seismotectonic atlas of Greece (v1.0) and its implementation. *Geosciences, 10*(11), 447. https://doi.org/10.3390/geosciences10110447

Kissing, E., Ellsworth, W., Eberhart-Phillips, D., & Kradolfer, U. (1994). Initial reference models in local earthquake tomography. *Journal of Geophysical Research, 99*(B10), 19635–19646. https://doi.org/10.1029/93JB03138
Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2021). An end-to-end earthquake detection method for joint phase picking and...

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks.

Zhu, W., Mousavi, S. M., & Beroza, G. (2020). Seismic signal augmentation to improve generalization of deep neural networks.

Zhu, W., & Beroza, G. C. (2019). Phasenet: A deep-neural-network-based seismic arrival-time picking method.

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep learning for seismic phase detection and picking in the after...