

RED-PAN: Real-Time Earthquake Detection and Phase-Picking With Multitask Attention Network

Wu-Yu Liao, En-Jui Lee[✉], Da-Yi Chen, Po Chen, Dawei Mu, and Yih-Min Wu

Abstract—In this article, we show that the real-time earthquake detection and phase picking with multitask attention network (RED-PAN) can carry out earthquake detection and seismic phase picking on real-time and continuous data with appropriate data augmentation. Goal-oriented data augmentations materialize the capability of RED-PAN. Mosaic waveform augmentation (MWA) synthesizes data conditioned by superimposed earthquake waveforms, marching MWA (MMWA) extends MWA to allow the dynamic input of seismograms, and earthquake early warning augmentation (EEWA) enables to identify P arrivals using the early part of P -wave waveforms. For stable P and S arrival probability distribution functions (pdfs) of continuous recordings, we use the median values of phase predictions at each time point until the model scans through, which we term the seismogram-tracking median filter (STMF). For real-time P arrival detection, we use a threshold (0.3) on the real-time P arrival pdf as the trigger criterion. We examined our proposed strategy in different application scenarios. For the dataset of the fixed-length samples, our RED-PAN(60 s) model performs similar to EQTransformer (EqT) on the STanford EArthquake Dataset (STEAD) and outperforms the Taiwan dataset. For continuous data examination of the 2019 Ridgecrest earthquake sequence, the number of earthquake waveforms detected by our RED-PAN(60 s) model is 2.7 times the number of EqT under the same receptive field (60-s-long seismogram). In the application of earthquake early warning (EEW), our RED-PAN(60 s) model only requires the P -wave waveform about 0.13 s long from the P -alert and 0.09 s long from the Taiwan Strong Motion Instrumentation Program (TSMIP) network. The source code is available at <https://github.com/tso1257771/RED-PAN>.

Index Terms—Data augmentation, multitask learning (MTL), real-time earthquake monitoring.

Manuscript received 1 January 2022; revised 6 May 2022 and 1 August 2022; accepted 28 August 2022. Date of publication 8 September 2022; date of current version 27 September 2022. The work of Wu-Yu Liao was supported in part by the Science College of National Cheng Kung University (NCKU Science); and in part by the Ministry of Science and Technology (MOST), Taiwan, under Contract MOST 109-2116-M-006-016. The work of En-Jui Lee was supported by the Ministry of Science and Technology, China, under Contract MOST 109-2116-M-006-016. The work of Po Chen was supported in part by the Nielson Energy Fellowship provided by the School of Energy Resources, University of Wyoming. This work was supported in part by the National Science Foundation's Major Research Instrumentation program under Grant 1725729 and in part by the University of Illinois at Urbana-Champaign. (Corresponding author: En-Jui Lee.)

Wu-Yu Liao and En-Jui Lee are with the Department of Earth Sciences, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: rickli92@gmail.com).

Da-Yi Chen is with the Central Weather Bureau, Taipei 10048, Taiwan.

Po Chen is with the Department of Geology and Geophysics, University of Wyoming, Laramie, WY 82071 USA.

Dawei Mu is with the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA.

Yih-Min Wu is with the Department of Geosciences, National Taiwan University, Taipei 10617, Taiwan.

Digital Object Identifier 10.1109/TGRS.2022.3205558

I. INTRODUCTION

EARTHQUAKE monitoring is the foundation of observational seismology, which is data-driven and aims to explore the dynamics and mechanisms of earthquakes. For short-term earthquake monitoring, earthquake early warning (EEW) systems can warn of impending ground shaking as an earthquake strikes [1], [2], [3]. The principle of EEW is to forecast seismic ground motions and disseminate alerts to different end users to take precautionary actions. According to the range of early warning areas, EEWs can be categorized as onsite warnings that issue alerts locally and regional warnings that issue alerts across a broader range of areas away from the hypocenter. Both the warning categories are based on observations of earthquake ground motions.

For regional warning, the source-based algorithms are one of the most popular choices. It relies on the observations of earthquake waveforms and the corresponding seismic phases recorded at stations over the seismic network [4], [5], [6], [7], [8]. For instance, ElarmS [4] aggregates P arrival information from several stations to determine source parameters, such as earthquake location, origin time, and magnitude. Then the spatial distributions of peak ground motions could be forecast using attenuation functions or ground motion prediction equations (GMPEs). The earthquake location and origin time are estimated using the seismic velocity models and seismic phase arrivals. Determination of magnitude using only a few seconds of P -wave data could be estimations of P -wave predominant frequency contents [1], [4] or predetermined logistic regressions [9].

Similarly, onsite warning depends on early detection of P -waves and forecasting impending peak ground motions *in situ* using single-station methods [10], [11], [12]. Hence, the earlier the P arrivals are detected, the faster the P -wave-based EEW algorithm kicks off. The quality of the P arrival picking determines whether the algorithms work as anticipated.

Due to the computational efficiency of the short-term average (STA)/long-term average (LTA) methods [13], they usually serve as defaulting real-time P arrival pickers. However, similar to other statistical algorithms that capture abrupt temporal variations in amplitude or pattern (e.g., earthquake envelope functions [14], autoregressive Akaike information criterion (AR-AIC) [15], kurtosis function [16], skewness function [17], [18], filtering [19], and particle motion polarization [20], [21], [22]), these traditional pickers rely on heavy parameter tuning and are error-prone under complex conditions, such as intense

aftershocks [23], [24], [25] and P -waves with very small amplitudes [23].

The deep-learning-based methods have recently significantly improved P -wave detection in quality and quantity using only a few seconds of P -wave data [26], [27]. More advanced deep-learning-based methods use full waveforms to detect earthquakes and/or pick both seismic P and S arrivals at the same time [28], [29], [30], [31], [32], [33], [34], which has the potential to provide more information for both the purposes of long-term earthquake monitoring and cataloging earthquakes for near-real-time applications [35], [36], [37].

The robustness and effectiveness of the deep-learning-based methods [29], [30], [31], [32], [33], [34] have made them emerging as baseline methods for processing the growing volume of continuous and real-time data. However, we have seldom found recent deep-learning-based models trained on datasets specific to real-time and continuous data, in which earthquake waveforms might lack complete representations of seismic sources. Here, we term such an incomplete earthquake waveform as a “time-clipped” earthquake waveform. The model with fixed-length input deals with time-clipped earthquake waveforms frequently when applied to real-time/continuous data. A common strategy to determine the model output values on continuous data is to track the results of repeatedly predicted samples for only a few seconds. We must note that the deep learning algorithms introduce a considerable amount of nonlinearity such that models could have different interpretations according to input data completeness. Our perspective is to track the model output until the model scans through all the data to achieve stable and objective output.

In this study, we propose a recurrent residual U-Net-based multitask attention network for earthquake waveform detection and seismic P/S arrival picking, in which a recurrent residual U-Net serves as a feature-sharing backbone network, and two attention networks tailored with different task learning. In addition, we present goal-oriented data augmentation techniques to detect P arrivals in real-time and stably process continuous data with conditions: successive events, time-clipped earthquake waveforms, and others. Our findings point out common problems overlooked by many studies: the training data are not specific to their usages, and their applications to continuous data were not model-objective. We promote model robustness on continuous and real-time data processing by adopting appropriate data augmentation techniques, which contribute a notable point of view to deep learning applications in observational seismology.

II. METHODS

Unlike the single-task seismic phase picking model of [34], in this study, we extend it to the multitask model, in which the attention gates construct the task-specific attention networks. In addition, we also use different data augmentation techniques on the training datasets and demonstrate performance improvements on continuous data and real-time detection. This section introduces the model architecture, the dataset in use,

the goal-oriented data augmentation method, and how we process continuous data. The goal-oriented data augmentation strategies materialize the capabilities of real-time earthquake detection and phase picking with multitask attention network (RED-PAN): 1) mosaic waveform augmentation (MWA); 2) marching MWA (MMWA); and 3) EEW augmentation (EEWA). MWA aims to pick and associate seismic P and S phases of independent earthquake waveforms on seismograms with multiple earthquake waveforms existing under the prediction window. MMWA is an extension of MWA that shifts the data backwardly and forwardly, helping to stabilize the model output value when making sliding predictions. EEWA allows early detection of seismic P arrivals for EEW. Considering the fixed length of the RED-PAN input and output, we adopt the seismogram-tracking median filter (STMF) strategy that tracks the prediction results and takes the corresponding median values of each sample when making sliding predictions on dynamic input continuous data.

A. Recurrent Residual U-Net-Based Multitask Attention Network

Multitask Learning (MTL) refers to the model learning fashion that enables the execution of multiple tasks in a single network by sharing global representations between different tasks [38], [39]. Generally, MTL has multiple loss functions to optimize and is widely used in the field of computer vision [40], [41] and natural language processing [42]. The MTL fashion is usually more desirable and efficient in practical usage in terms of memory, inference speed, and model setup. MTL poses two critical challenges to sharing informative features among different tasks: 1) how to share, namely, model architecture design; and 2) how to balance loss from different tasks. Under the framework of a multitask attention network [43], our adaptive proposal RED-PAN is composed of a recurrent residual U-Net [34], [44] as the backbone network for global feature-sharing, and two mounted attention subnetworks for task-specific feature learning (Fig. 1). Each attention network is computed upon several attention modules in an encoder–decoder manner, which has the same depth as the recurrent residual U-Net. The attention modules apply soft attention gates [43] on the corresponding recurrent-residual convolution (RRC) layers [45] of shared recurrent-residual U-Net at all the levels [Fig. 1(a) and Appendix A], which serve as feature selectors tailored with each task. The model outputs phase time functions for picking seismic phases and earthquake waveform detection functions from the last layer of the two attention subnetworks separately. Both the attention subnetworks correlate with the backbone network, and the total loss from the weighted loss of attention subnetworks would be minimized after training. Thus, the shared recurrent residual U-Net could learn generalized representations across all the tasks, and each attention network could jointly learn task-specific features in a self-supervised manner.

The input of RED-PAN(60 s) is the Z score standardized raw three-component seismograms with a length of 6000 samples, which corresponds to 60-s-long seismograms with a

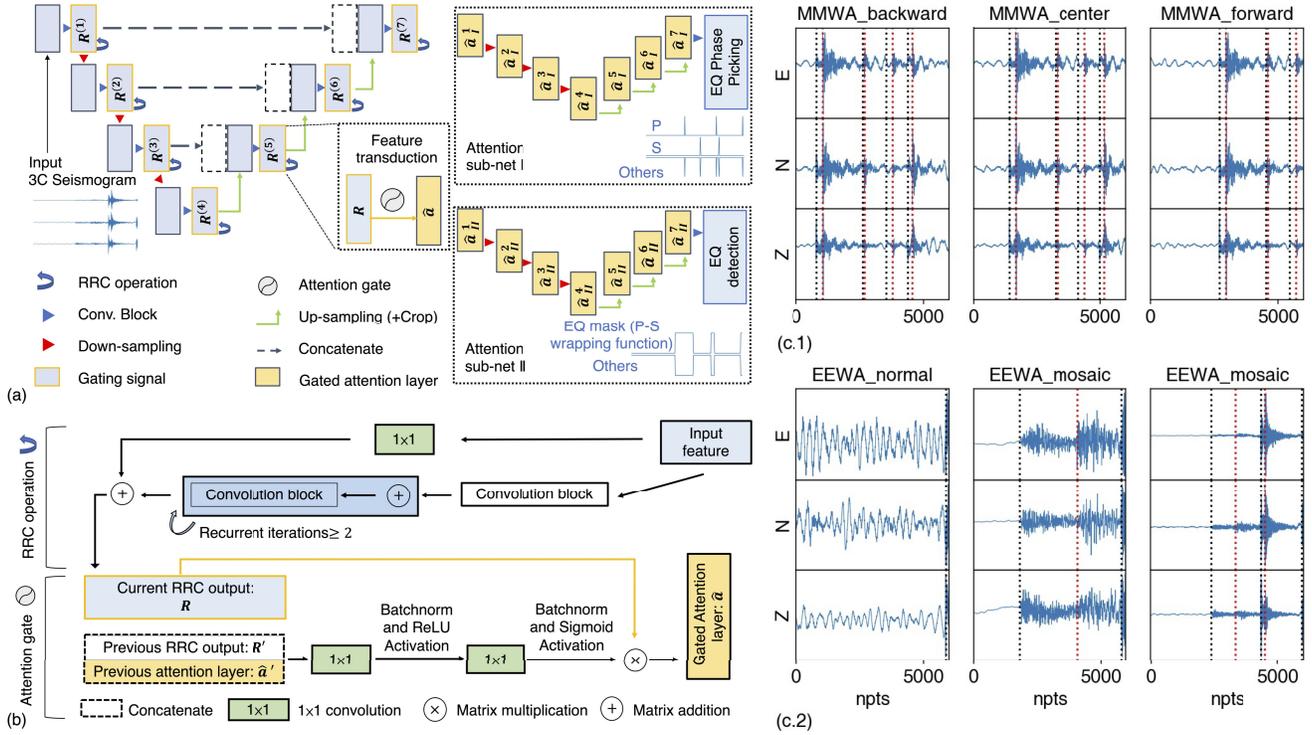


Fig. 1. Multitask model architecture and goal-oriented data augmentation templates for seismic phase picking and earthquake event detection. (a) Model consists of a recurrent-residual U-Net as the backbone and two mounted task-specific attention networks that share information in-between the backbone. The model inputs are the three-component seismograms, and the attention networks of the model output seismic phase picking functions (attention sub-net I) and earthquake waveform detection functions that wrap every P - S pair (attention sub-net II). Here, RRC operation refers to RRC operations. Each attention sub-net is composed of several gated attention layers $\hat{\mathbf{a}}$ corresponding to the gating signal, or RRC outputs \mathbf{R} of recurrent residual U-Net at all the depths. Feature transduction between $\hat{\mathbf{a}}$ and \mathbf{R} relies on the attention gate mechanism. (b) Details of the RRC operation and gated attention layer computation for constructing the attention network. (c.1) Triplet set of MMWA products that show the backwardly and forwardly shifting of the centered quasi-synthetic waveform. (c.2) Templates of EEWA products with the black and red dashed lines denoting P and S phase arrivals, respectively. The first template shows the waveform of the first few seconds of the P -wave recorded under background noise interference. The second and third quasi-synthetic templates, respectively, show the waveform that the first few seconds of the P waves recorded under the interference of another/the other two earthquake waveforms.

100-Hz sample rate. The outputs are two sets of vectors for seismic phase picking and earthquake waveform detection (Figs. 1 and 7). The target functions for the seismic phase picking module are composed of three phase time functions. The first two functions are P and S phase time functions, which are the truncated Gaussian functions with the standard deviation of 0.2 and 0.3 s centered at labeled P and S arrivals. Other samples are padded with 0 to fulfill the length of 6000 point-to-point samples. The last function of seismic phase picking module is “Others” that the probability distribution function (pdf) is calculated as: $\text{target}(\text{Others}) = 1 - \text{target}(\text{P}) - \text{target}(\text{S})$. For every time step, the summation of the three-channel target functions is 1, so we could apply softmax normalized exponential function to set probabilities in the output layer for the phase picking module. Note that the standard deviation for the labeled phase inherently represents the labeling misfits of the ground truth. For model calibration, we have conducted experiments to find the optimal standard deviation of the truncated Gaussian function of the P and S phases of the Taiwan dataset [34]. The target functions for the earthquake waveform detection module are composed of two pdfs: 1) the box-car-like function wrapping a P and S arrival pair [abbreviated as “EQ mask” in Figs. 1(a) and 7] and 2) “Others” computed as: $\text{target}(\text{Others}) = 1 - \text{target}(\text{EQmask})$. The data space of “EQ mask” between the ground truth P and S arrivals

is filled with 1. The front is replaced by the P phase time function (half truncated Gaussian function) before ground-truth P arrival, and the same steps are used for the tail of the box car using the S phase time function (Fig. 7). Also, we apply the softmax normalized exponential function to set probabilities in the output layer for the earthquake waveform detection module. The loss function \mathbf{H} for the two tasks is defined in terms of cross-entropy between the softmax normalized predicted functions \mathbf{q} and the corresponding ground-truth target functions \mathbf{p} as follows:

$$\mathbf{H}_{\text{picking}}(\mathbf{p}, \mathbf{q}) = - \sum_c^3 \sum_x \mathbf{p}_c(x) \log \mathbf{q}_c(x) \quad (1)$$

$$\mathbf{H}_{\text{EQmask}}(\mathbf{p}, \mathbf{q}) = - \sum_c^2 \sum_x \mathbf{p}_c(x) \log \mathbf{q}_c(x). \quad (2)$$

In (1) and (2), x refers to the time step and c is the number of output functions for each task.

B. Balancing Loss From Different Tasks

Generally, the loss function for \mathbf{k} -task learning is defined as the weighted sum of different task losses. Training MTL models have difficulty striking a balance between different tasks, which is particularly tedious for manual tuning. Grad-Norm [46] learns to average task weighting over time by

considering the rate of change in loss for different tasks, which requires access to the internal gradients of the model. Similarly, here we adopt a dynamic weight average (DWA) [43] strategy that considers only numerical task loss. Let \mathbf{L}_k denote the loss function for task k . The task-specific weighting λ_k is defined as follows:

$$\lambda_k(I) := \frac{K \times \exp(w_k(I-1)/H)}{\sum_i \exp(w_i(I-1)/H)}, \quad w_k(I-1) = \frac{\mathbf{L}_k(I-1)}{\mathbf{L}_k(I-2)}. \quad (3)$$

Here, I represents training epoch iterations, and we set $w_k(I) = 1$ for $I \leq 2$. $w_k(\cdot)$ calculates the relative descending rate, which is then scaled by the temperature scaling factor H . A larger H results in more even distribution weightings among different tasks, and we set $H = 2$ in this study. Furthermore, (3) can be considered the multiplication of the softmax operator and constant K , which ensures that $\sum_i \lambda_i(I) = K$. Therefore, in this study the total loss in the training epoch I could be calculated as follows: $\sum_k \lambda_k(I) \mathbf{H}_k$, where $k = 1, 2$ represents the phase picking task and the earthquake waveform detection task.

C. Dataset

In this study, we collect seismograms sampled at 100 Hz from the Taiwan dataset and STanford EArthquake Dataset (STEAD) [47] for model training, validation, and testing. Seismograms of the Taiwan dataset are recorded by the Central Weather Bureau (CWB) network and Broadband Array in Taiwan for Seismology (BATS) [48], with data gathered from 2012 to 2018 used for model training and validation and those gathered from 2019 used for model testing. To ensure that the model can characterize earthquake waveforms at any possible location within the prediction window, we slice and randomly locate the earthquake waveform of the Taiwan dataset for each sample. Different from our Taiwan dataset, the earthquake waveform of STEAD is located at the forepart of the seismogram so that a complete earthquake waveform can be well-presented. On the other hand, to meet the needs of single-component seismic recorders and recording interruptions, we randomly padded the waveform with zeros by fractions or dropped some channels. The STEAD dataset used for model training and testing in this study is the same as that of [33], which leaves 880 K earthquake waveforms and 200 K non-earthquake waveforms for model training and 103 K earthquake waveforms and 23.5 K non-earthquake waveforms for benchmarking. The Taiwan testing dataset is composed of 198 K earthquake waveform samples and 155 K non-earthquake samples. For the 198 K earthquake waveform samples, we create two datasets, with and without ground-truth P arrivals fixed at the third second to ensure the completeness of the earthquake waveform. The earthquake sample without fixed P locations has ground-truth P arrival randomly distributed, which satisfies the rules of making a single-event dataset described in Appendix B. We wonder whether the model can perform similarly on these two datasets to account for the model robustness toward the same dataset with different earthquake waveform locations.

Our proposed RED-PAN is trained/validated on 500/100 K samples, with a composition of 39.5% Taiwan single-event earthquake waveform, 12% MMWA products of Taiwan data, 8% EEWA products of Taiwan data, 8.25% non-earthquake waveform of Taiwan data, 12% STEAD single-event earthquake waveform, 12% MWA products of STEAD, and 8.25% non-earthquake waveform of STEAD.

D. Goal-Oriented Data Augmentation

1) *Mosaic Waveform Augmentation (MWA)*: The core idea of MWA comes from the fact that earthquakes are triggered frequently in a short time in regions of active seismicities, such as Taiwan, southern California, and Japan. Thus, it is common that several earthquake waveforms, including the time-clipped earthquake waveform and superimposed earthquake waveform, are concurrently visible under the prediction window of models. To consider such conditions, MWA generates semisynthetic earthquake waveforms by randomly superimposing two to four different/identical scaled earthquake waveforms collected from the same recorder, namely, the ‘‘mosaic waveform.’’

2) *Marching MWA (MMWA)*: As an extension of MWA, MMWA further randomly shifts the waveform backwardly and forwardly (which is the meaning of ‘‘marching’’) to form a triplet set: ω_{backward} , ω_{center} , and ω_{forward} [Fig. 1(c.1)]. Every earthquake waveform in ω_{center} is paired with P and S arrivals, while those in ω_{backward} and ω_{forward} may be time-clipped due to random shifts.

3) *EEW Augmentation (EEWA)*: Fast and accurate P arrivals of large earthquakes on real-time seismograms enable quicker activation of EEW systems and provide more reliable source parameter estimates. EEWA generates earthquake waveforms with only P arrivals available for rapid P wave detections, which might also be similar to some MMWA products. While EEWA products include P phase waveforms with a broader range of source–receiver distance and various patterns of P phase features. All EEWA products include an earthquake waveform of only the P phase available at the end of the seismogram, with or without other earthquake waveforms existing in the forepart [Fig. 1(c.2)]. This ensures that the model could learn to characterize P arrivals under background noise and interference of other seismic waves.

E. Seismogram-Tracking Median Filter (STMF)

In the scenario of making sliding predictions on continuous data, the length of the model input, completeness of earthquake recordings, shape, and position under the prediction window may influence the prediction results even for the same earthquake waveform. Among them, the completeness of earthquake recordings and shape is directly related to the source–receiver distance, source mechanisms, recorder responses, and background noise. Also, due to the dynamic input characteristics of sliding predictions, the model will make predictions on time-clipped earthquake waveforms. The longer the earthquake waveform, the less time the entire earthquake waveform can be viewed by the model, while the length of the model input determines the maximum length

TABLE I
STATIC PREDICTION PERFORMANCE ON THE STEAD AND TAIWAN DATASET

Dataset	Type	Model	TruePositive (%)	Mean (s)	std (s)	Precision	Recall	F1 score	MAE (s)	
STEAD	P	RED-PAN(60s)	0.9822	-0.0160	0.0790	0.9890	0.9947	0.9919	0.0435	
		EqT	0.9894	-0.0034	0.0297	0.9961	0.9941	0.9951	0.0110	
	S	RED-PAN(60s)	0.9541	-0.0108	0.1199	0.9670	0.9905	0.9786	0.0764	
		EqT	0.9389	-0.0196	0.1051	0.9810	0.9649	0.9729	0.0662	
	Mask	RED-PAN(60s)	0.9967	-			0.9945	0.9967	0.9956	-
		EqT	0.9993				0.9997	0.9993	0.9995	
TW	P	RED-PAN(60s)	0.9496	0.0029	0.0778	0.9822	0.9747	0.9784	0.0414	
		EqT	0.8037	0.0255	0.1073	0.9690	0.8402	0.9000	0.0636	
	S	RED-PAN(60s)	0.8900	0.0083	0.1317	0.9490	0.9539	0.9514	0.0855	
		EqT	0.6877	0.0447	0.1114	0.9726	0.7269	0.8320	0.0804	
	Mask	RED-PAN(60s)	0.9854	-			0.9931	0.9854	0.9892	-
		EqT	0.8744				0.9979	0.8744	0.9321	
TW (fixP)	P	RED-PAN(60s)	0.9522	-0.0001	0.0778	0.9831	0.9761	0.9796	0.0422	
		EqT	0.7583	-0.0271	0.0373	0.9763	0.7865	0.8712	0.0334	
	S	RED-PAN(60s)	0.8884	0.0086	0.1318	0.9491	0.9523	0.9507	0.0859	
		EqT	0.6962	0.0465	0.1121	0.9707	0.7345	0.8362	0.0815	
	Mask	RED-PAN(60s)	0.9914	-			0.9932	0.9914	0.9923	-
		EqT	0.8830				0.9979	0.8830	0.9370	

fixP refers to the fixed P arrival location at the 300-th point of the sample; the bold style indicates better performance between RED-PAN(60s) and EqT; the red text points out the obvious performance discrepancy of EqT on Taiwan testing dataset.

of the entire earthquake waveform it can accommodate. The deep-learning-based model can have different phase arrival interpretations according to waveform completeness under the prediction window. Therefore, considering the physical restrictions between the model input length and waveform length, we determine the output pdfs by storing the corresponding model output values of repeatedly predicted data samples and taking the median value as the STMF output as soon as the prediction window slides across all the samples. By doing so, we ensure that the STMF output is relatively objective than simply tracking the peak values for a few seconds, which the model might misjudge owing to the incompleteness of the time-clipped earthquake waveform.

III. RESULTS

We evaluated our model under different application scenarios and compared it with EQTransformer (EqT) [33], which is one of the most advanced models.

1) For earthquake waveforms located under the model prediction window, such as those encountered during phase pick revision [24], we benchmark phase picking and earthquake waveform detection performance on the Taiwan dataset and STEAD [47].

2) To evaluate the phase picking and earthquake waveform detection performance on continuous data, we benchmark the model on the 2019 Ridgecrest M_w 7.1 earthquake sequence with a ground-truth catalog [24] developed by template matching algorithm (TMA) [49].

3) For the (near) real-time P arrival detection, we evaluated models on event-based seismograms collected from two dense seismic networks in Taiwan, P -alert [50] and Taiwan Strong Motion Instrumentation Program (TSMIP) [51]. We wonder how much data are required for models to trigger P arrivals using the limited length of earthquake waveform and how accurate the P arrival triggers are.

A. Static Prediction Performance

In this section, we evaluate the model performance on the test dataset of STEAD and the Taiwan dataset in the length

of 6000 samples (Table I). We compare our RED-PAN(60 s) model with EQTransformer (here, we denote it as “EqT” [33]), which has the maximum separation between P and S arrivals of 49.53 s in the training data. The raw waveform and waveform bandpassed at 1–45 Hz are the inputs for RED-PAN(60 s) and EqT, respectively. The threshold for true picks is the 0.5-s-long absolute time difference between the predicted one and the ground truth; the picks with peak values larger than 0.3 are counted as positive picks. Here, we use the threshold value of 0.3 used by EqT [33] for fair comparisons. For detection mask evaluation, the confusion matrix is constructed on the detection results of earthquake and non-earthquake samples. We consider a true positive if the mean value of mask function wrapping P and S arrivals of an earthquake is larger than 0.5, while the sample would be characterized as a non-earthquake sample if the same criterion cannot be met. As shown in Table I, RED-PAN(60 s) and EqT perform similarly on STEAD, but RED-PAN(60 s) generally outperforms EqT on the Taiwan dataset. On the other hand, considering the prediction results of the Taiwan test data with or without a fixed P location, EqT performance differentiates. While RED-PAN(60 s) performs quite similarly, indicating more stable prediction results for the same earthquake waveform with different locations under the same receptive field.

B. Continuous Data Examination

To achieve objective evaluations on continuous data processing, we apply the RED-PAN models and EqT with the STMF strategy on continuous data gathered from 22 stations located in the Ridgecrest region, southern California, from 4 July, 2019 16:00 to 8 July, 2019 00:00 (UTC + 0). The ground truth of the earthquake event catalog with magnitudes larger than 0 is obtained from TMA described in [24], which left us 34 381 event templates. The location of event templates lies in the longitudinal range from -118.1215 to -117.2421 and the latitudinal range from 35.4991 to 36.2495 . The sliding prediction window is 4 s in this test.

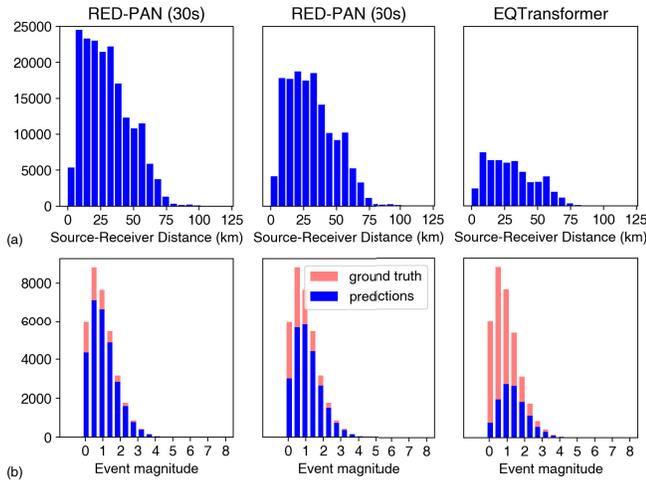


Fig. 2. Continuous data examination results of 22 stations in the Ridgecrest region, southern California, from 4 July, 2019 16:00 to 8 July, 2019 00:00 (UTC + 0). (a) Source-receiver distance histogram of valid event waveform detection of all the stations. (b) Event magnitude histogram of ever detected events.

Valid event waveform detection is counted as the prediction results fulfill the following statements: 1) both the P and S pick probabilities are larger than 0.3; 2) the mean value of the detection mask between the P and S picks is larger than 0.5; and 3) the positions of predicted P - S pairs should lie within $[\alpha_p^{\psi,Z} - 1.5 \text{ s}, \alpha_p^{\psi,Z} + 1.5 \text{ s}]$, where $\alpha_p^{\psi,Z}$ is the labeled phase arrivals that are forward-estimated by seismic velocity, model, F3DT [52], and source-receiver distance between the hypocenter ψ and the station Z .

Restricted to the data preprocessing step of Z score standardization, an event with a large amplitude under the prediction window would compress other signals. The longer the model can accommodate, the more data space would be compressed. Hence, event waveforms of small amplitude might approximate the background noise level if large-amplitude events are recorded close in time, making them undetectable. We compare the performance of EqT and our proposed RED-PAN models with input lengths of 30 and 60 s, denoted as RED-PAN(30 s) and RED-PAN(60 s), respectively, (Fig. 2). The way RED-PAN(30 s) is trained is identical to that of RED-PAN(60 s) while having the maximum separation between P and S arrivals of 25 s in the training data.

In summary, EqT detects 12283 events (36%) with 55548 valid event waveform detections at all the stations; RED-PAN(60 s) detects 24339 events (71%) with 153472 valid event waveform detections, and RED-PAN(30 s) detects 28824 events (84%) with 189794 valid event waveform detection. Considering valid event waveform detection at all the stations, RED-PAN models outperform EqT across all the ranges of source-receiver distances [Fig. 2(a)]. RED-PAN(60 s) has 2.7 times more valid event waveform detection than EqT under the same receptive field. RED-PAN(30 s) detects more earthquake waveforms with small magnitudes and short source-receiver distances, which might appear as relatively small-amplitude waveforms. Such results may verify the argument that earthquake waveforms with smaller amplitudes are

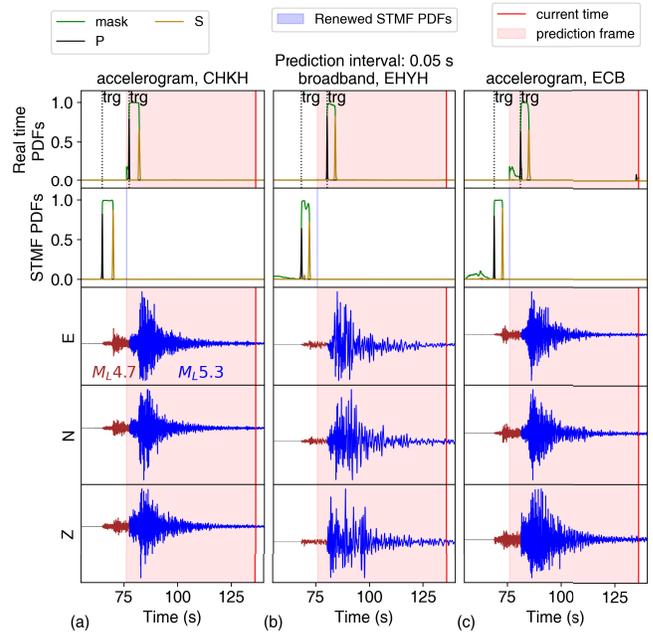


Fig. 3. Examples of real-time data processing using RED-PAN(60 s) model on raw seismograms starting from 24 March, 2022 00:37:10.00 (UTC + 0), recorded at the three stations, (a) CHKH, (b) EHYH, and (c) ECB, closest to the epicenters. The seismograms recorded an M_L 4.7 earthquake followed by an M_L 5.3 earthquake that took place very close in time and space, causing the waveform of the latter one to superimpose on the former one. The first row shows real-time prediction pdfs, and the second shows the pdfs obtained with STMF processing. We renew the STMF pdfs when the prediction window moves on to the next time stamp. Thus, the length of renewed pdfs is equal to the prediction interval. The red frame represents the prediction window with a 0.05-s sliding interval. We perform peak detection on the P phase time function at each prediction time step to determine P arrival triggers. The trigger is issued if the peak value is larger than 0.3 existing within the first second of the incoming data, e.g., in the range of $[\mathbf{t}_{\text{current}} - \mathbf{1}, \mathbf{t}_{\text{current}}]$. The dotted black lines label the P phase triggers.

likely to be missed by a model with a longer input length if large-amplitude events exist closely in time.

C. Efficiency and Accuracy of P Arrivals Triggering

The efficiency of picking P arrivals of earthquakes urgently for EEW depends on two key factors: 1) the inference speeds of the algorithms, which are 0.0150 and 0.0261 s for RED-PAN(30 s) and RED-PAN(60 s), respectively, using an Intel (R) Xeon (R) W-2125 CPU at 4.00 GHz in 1000 times, on average; and 2) how long the earthquake waveform is required for algorithms to pick P arrivals.

Fig. 3 demonstrates a scenario template showing how RED-PAN(60 s) processes the incoming data, with the 0.05-s interval of sliding prediction on the three-component seismograms sampled at 100 Hz. The seismograms in Fig. 3 show an M_L 4.7 earthquake waveform superimposed by an M_L 5.3 earthquake waveform that occurred in east Taiwan (Fig. 4). The M_L 5.3 earthquake has met the criteria for issuing earthquake warnings, but the current system has missed the event. The first row shows the pdfs of the model current output, spanning from $(\mathbf{t}_{\text{current}} - 60) \text{ s}$ to the current time $\mathbf{t}_{\text{current}}$, represented by the red frame, which can provide information for real-time applications. For real-time detection of the P -wave, we perform peak detection on the current output of the P

TABLE II
P ARRIVAL PICKING EFFICIENCY ON *P*-ALERT AND TSMIP DATA

		RED-PAN(30s)	RED-PAN(60s)	MMWA model(60s)	MWA model(60s)	STA/LTA(CWB)
P-alert	Recall	0.9840	0.9796	0.8278	0.2253	0.6006
	Average length of waveform to trigger (s)	0.1110	0.1294	0.2087	0.7423	-
	MEAN($t_{\text{urg}} - t_{\text{STMF}}$) (s)	-0.0974	-0.0670	-0.0244	-0.0581	-
TSMIP	Recall	0.9982	0.9967	0.9717	0.5113	0.8044
	Average length of waveform to trigger (s)	0.0787	0.0853	0.1148	0.6415	-
	MEAN($t_{\text{urg}} - t_{\text{STMF}}$) (s)	-0.1030	-0.0770	-0.0577	-0.0503	-

In this study, a true positive trigger is counted when *P* arrival is detected within a ± 1 second window around the ground-truth with probability larger than 0.3, and lies in the range of $[t_{\text{current}} - 1, t_{\text{current}}]$ in the prediction window. The sliding prediction window for making this table is 0.05 seconds. t_{urg} is the first trigger of *P* arrivals, and t_{STMF} is the *P* arrivals obtained with STMF strategy.

phase time function and issue a trigger if its peak value is larger than 0.3 within the first second of the incoming data, e.g., in the range of $[t_{\text{current}} - 1, t_{\text{current}}]$. The second row is the STMF pdfs obtained with STMF processing, with the length of the renewed data equal to the model sliding interval. When the model moves on to the next time stamp with a sliding interval, it will no longer process the previously monitored waveform in equal length, and then the corresponding STMF pdfs of these data could be renewed. In Fig. 3, the peaks (i.e., triggers) detected within the range of $[t_{\text{current}} - 1, t_{\text{current}}]$ with value larger 0.3 from model output *P* phase time function are labeled with dotted lines. We also provide an animation of Fig. 3: <https://youtu.be/582yB1zigWE>. In this example, our PED-PAN model can accurately detect the *P*-waves in (near) real-time of the $M_L 5.3$ earthquake to avoid missing the event in the current EEW systems.

To evaluate the shortest amount of waveform that RED-PAN requires to characterize *P* arrivals, we additionally collected seismograms recorded by two dense strong motion seismic networks in Taiwan, *P*-alert [50] and the TSMIP [51]. This study's model training data do not include all TSMIP and *P*-alert data. We used the top ten nearest station recordings from 163 events of the *P*-alert network and 274 events of the TSMIP network (Fig. 4) from 2013 to 2019. The magnitude of events ranges from 4.0 to 6.91, and most of the seismograms are collected from stations with source–receiver distances less than 25 km. We compare the performance between the STA/LTA algorithm tuned by CWB experts, RED-PAN(30 s), RED-PAN(60 s), and models separately trained using the MMWA and MWA strategy only. The true positive trigger is counted when the *P* arrival is detected within a ± 1 s window around the ground truth with probability larger than 0.3 and lies in the range of $[t_{\text{current}} - 1, t_{\text{current}}]$ under the prediction window. The false negative trigger is counted if no trigger is detected. We compute the true positive rate (or recall rate) to estimate the model performance on the task of triggering *P* arrivals. In this test, we made sliding predictions with the interval of 0.05 s on the collected TSMIP and *P*-alert data. For online applications, the decision of prediction interval must also consider the model inference time on the installed machine, the algorithms' buffering, data flow, and other background programs.

Fig. 5 shows the recall rate of *P* arrivals triggering across different source–receiver distances and reaction times (i.e., length of waveform needed) to trigger *P* arrivals using three-component/vertical component data. The histograms of the

true positive *P* triggers show that the RED-PAN models outperform all source–receiver distances and perform better on the TSMIP data than on the *P*-alert data. Table II lists the *P* arrival picking efficiency on the *P*-alert and TSMIP data. Generally, the RED-PAN models outperform both the *P*-alert and TSMIP network data in terms of recall rate and the average length of the waveform required for triggering. Among the RED-PAN models, the performance of RED-PAN(30 s) and RED-PAN(60 s) is comparable. On the other hand, an obvious performance gap exists between the TSMIP data and the *P*-alert data of all the compared methods, including the STA/LTA algorithm. We attribute this to the data quality difference between the traditional sensors and the sensors of micro-electro-mechanical systems (MEMS) that are composed of the *P*-alert network. As [53] concludes, by far, the strong self-noise and lower responses at a low frequency of MEMS compared with the traditional devices make part of the seismic background noise unrecognizable. Same conditions could also be observed in the *P*-alert data if the amplitude of ground motion is not large enough, making it require more data to recognize the seismic *P* phase. In addition, Fig. 6 demonstrates the triggering error comparison between RED-PAN(60 s) and the STA/LTA method tuned by CWB experts, where only time misfits less than 1 s are plotted since the STA/LTA method misses or overlooks a portion of *P* arrivals as shown in Table II. Fig. 6 shows that the misfits of *P* arrival triggers from both the RED-PAN(60 s) and STA/LTA methods lie in a ± 0.1 s interval, while picks of RED-PAN(60 s) are closer to the ground truth with a higher recall rate (Table II). In practical EEW applications, different criteria are used to avoid false positive picks leading to false alarms. For example, a triggered *P* wave needs checks with other criteria (e.g., amplitude, signal-to-noise ratio, and *P*-wave peak displacement for magnitude estimate) for a valid pick in the Taiwan EEW system [8]. In addition, a stable hypocenter derived from qualified *P* picks at different stations and the estimated ground shaking larger than the threshold are required before an EEW can be issued.

IV. DISCUSSIONS

A. Earthquake Waveform Pattern and Model Generalization

How earthquake waveform pattern influences the model performance is usually termed as model generalization in seismic phase picking and earthquake detection task, which is directly related to earthquake source parameters (i.e., focal

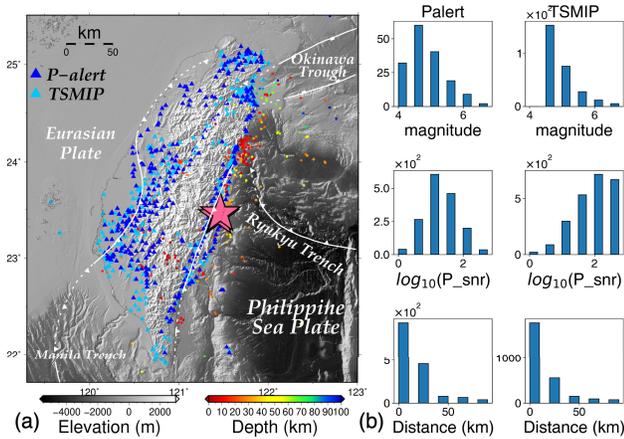


Fig. 4. Distribution of data used to evaluate P arrival picking efficiency. (a) Seismic map of used events and seismic networks. (b) Data distribution of the P-alert and TSMIP over magnitude, SNR, and source-receiver distance. The two pink stars denote the M_L 4.7 earthquake and M_L 5.3 earthquake that occurred on 2022-03-24 00:37 (UTC + 0).

mechanisms and depths), the medium which seismic waves propagate through, and recording environments. In Table I, 10%–20% true positive ratio gap of the static prediction performance exists between EqT and RED-PAN(60 s) on the Taiwan dataset, which can be illustrated by the precision and recall rate of phase picks. The precision rate is similar between RED-PAN(60 s) and EqT, but the recall rate of EqT is much lower than that of RED-PAN(60 s), indicating more false negatives. Such a result implies that both RED-PAN(60 s) and EqT could find the accurate and similar position of phase arrivals, while EqT is not that confident in the Taiwan dataset, suggesting that the training data specific to the region of interest are no less critical than the model architecture design.

B. Insights of Goal-Oriented Data Augmentation

In Table II, even when trained without EEWA, the MMWA model can trigger many P arrivals, which works as anticipated since the movement of “marching” would produce a time-clipped earthquake waveform that only contains the P phase. The performance of the MWA model, in which the training strategy is similar to [33], [54], could support the above argument. It can hardly pick P arrivals using time-clipped earthquake waveforms without marching waveforms in the training data. In addition, in Fig. 5, the performance discrepancy on P arrival triggering between the RED-PAN models and MMWA model lies in samples of more considerable source-receiver distances, where the performance of the MMWA model drops with increasing source-receiver distance. We attribute such results to the training data diversity of time-clipped earthquake waveforms containing only the P phase. MMWA products have a limited number of samples with longer source-receiver distances since they can hardly be accommodated within a 60-s window with other earthquake waveforms to compute mosaic waveforms. However, the computation of EEWA products is not limited to the source-receiver distance of the earthquake waveform so that RED-PAN can trigger P arrivals with a broader range of source-receiver distances.

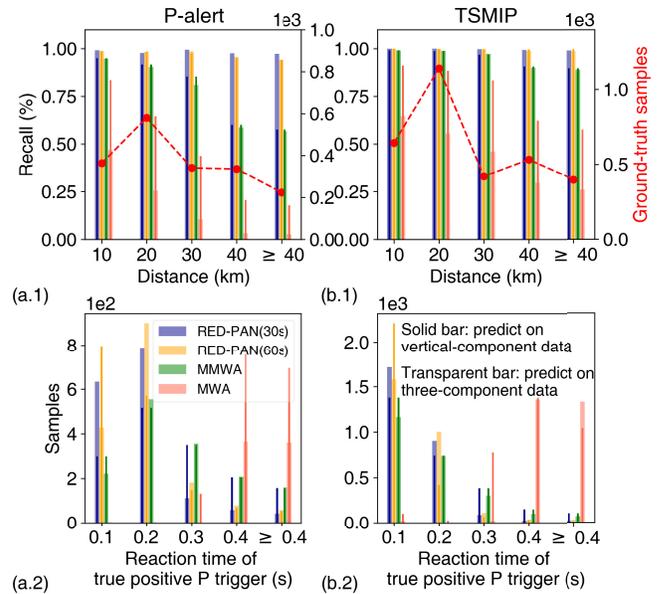


Fig. 5. Histograms of the recall rate of P arrivals triggering and triggering reaction time (i.e., length of waveform needed to trigger P arrivals) of (a.1) and (a.2) P-alert and (b.1) and (b.2) TSMIP networks. Four models are performed on both three-component and vertical component data, represented by transparent lines and solid lines, respectively. The MWA model and MMWA model separately indicate model trained with single-event earthquake waveform along with MWA and MMWA strategies. Generally, the RED-PAN models outperform across source-receiver distances, and most RED-PAN picks are triggered using less than 0.2 s of data.

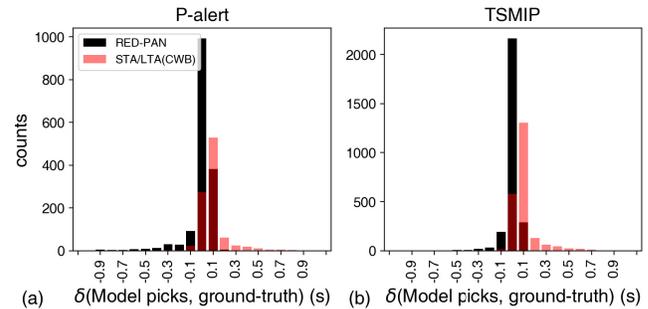


Fig. 6. Histograms of triggering error of RED-PAN(60 s) and STA/LTA methods tuned by CWB expert on (a) P-alert and (b) TSMIP network.

V. CONCLUSION

In this study, we trained our RED-PAN model with a recurrent residual U-Net-based multitask attention network in a self-supervised manner, which dynamically adjusts the weightings of seismic phase picking task and the task of computing earthquake detection mask during training. The core idea of this study is to leverage the goal-oriented data augmentation techniques, MWA, MMWA, and EEWA, which remarkably improve the model performance to achieve continuous and real-time data processing. Our research shows that the performance of the deep-learning-based models in different application scenarios, such as static prediction and continuous/real-time data, is highly related to the variations in training data. Aided by the rapid developments of deep learning studies, more advanced architectures have been applied to seismological problems with various input and target outputs. However, unlike computer vision or natural language processing, the

input relevant in our field, seismogram, is nonstationary. Our perspective of developing models with goal-oriented data augmentation may provide another trajectory to embark on.

APPENDIX A

FEATURES TRANSMISSION AMONG ATTENTION NETWORKS AND RECURRENT RESIDUAL U-NET

Let us denote $\mathbf{R}^{(j)}, j \geq 1$ as the shared features in the j th level RRC layer of the shared recurrent residual U-Net and $\hat{\mathbf{a}}_i^{(j)}$ as the learned attention gate (or mask) in the layer j for task i . The task-specific features $\hat{\mathbf{a}}_i^{(j)}$ are computed by elementwise multiplication of attention gates and shared features. As Fig. 1(b) shows, our attention modules take two inputs: the output features of the shared RRC layer $\mathbf{R}^{(j)}$; the concatenation of previous attention features $\hat{\mathbf{a}}_i^{(j-1)}$ and the shared RRC layer $\mathbf{R}^{(j-1)}$. Apart from the first attention module that takes only $\mathbf{R}^{(1)}$ as input features, other task-specific attention features computed from the encoder and the decoder are formulated as follows:

$$\hat{\mathbf{a}}_{i,\text{enc}}^{(j)} = h_i^{(j)} \left(g_i^{(j)} \left(f_i^{(j)} \left(\left[\hat{\mathbf{a}}_{i,\text{enc}}^{(j-1)}; \mathbf{R}^{(j-1)} \right] \right) \right) \right) \odot \mathbf{R}^{(j)}, \quad j \geq 2 \quad (4)$$

and

$$\hat{\mathbf{a}}_{i,\text{dec}}^{(j)} = g_i^{(j)} \left(f_i^{(j)} \left(\left[u_i^{(j)} \left(\hat{\mathbf{a}}_{i,\text{dec}}^{(j-1)} \right); \mathbf{R}^{(j-1)} \right] \right) \right) \odot \mathbf{R}^{(j)} \quad (5)$$

where \odot denotes the elementwise multiplication; $f_i^{(j)}$ and $g_i^{(j)}$ are the convolutional layers of $[1 \times 1]$ kernels with batch normalization, following ReLU and sigmoid activation, respectively; $h_i^{(j)}$ and $u_i^{(j)}$ represent the convolutional down-sampling and upsampling layers that enable matching of the corresponding resolution. For more details about RRC input and output formulations for seismograms, we refer to [34].

APPENDIX B

DETAILS OF SINGLE-EVENT DATASET

Apart from the complete earthquake waveform in STEAD, we also consider that the deep learning algorithms are capable of picking and associating seismic P and S arrivals with incomplete waveforms that might be informative enough. We define the length of basic informative earthquake waveform as $\gamma_{\text{basic}} = \mathbf{m} \times \alpha_{\text{res}} + \epsilon_P$, with \mathbf{m} times the absolute time residual between labeled P and S arrivals (α_P and α_S) in seconds: $\alpha_{\text{res}} = |\alpha_S - \alpha_P|$, $\alpha_{\text{res}} > 0$; and $\epsilon_P = 0.5$ s the potential P arrival picking error. Also, with $\gamma_{\text{bef-P}}$ and $\gamma_{\text{aft-basic}}$ defined as randomly distributed space before the point ($\alpha_P - \epsilon_P$) and after the endpoint of γ_{basic} , the length for each informative earthquake waveform sample γ_{total} in the Taiwan dataset can then be formulated as follows:

$$\gamma_{\text{total}} = \gamma_{\text{bef-P}} + \gamma_{\text{basic}} + \gamma_{\text{aft-basic}} \quad (6)$$

where \mathbf{m} is controlled by α_{res} as follows:

$$\mathbf{m} = \begin{cases} 2, & \text{if } \alpha_{\text{res}} < 20 \\ 1.5, & \text{if } 20 \leq \alpha_{\text{res}} < 25 \\ 1.2, & \text{if } 25 \leq \alpha_{\text{res}} < 40. \end{cases}$$

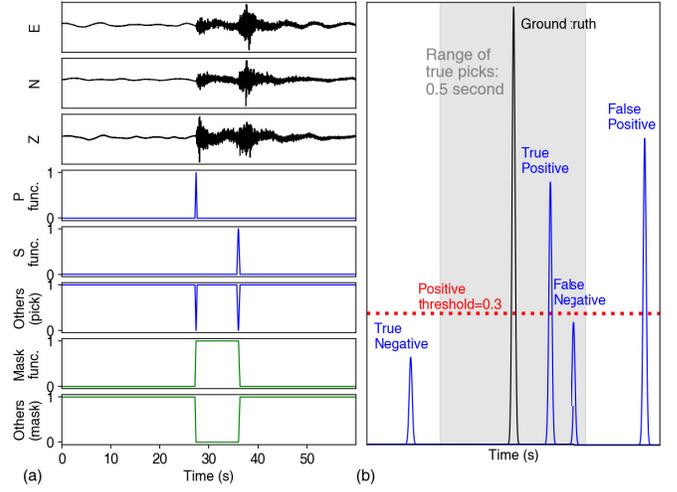


Fig. 7. RED-PAN input template. (a) Input seismogram and the corresponding target functions for the seismic phase picking modules (blue line) and earthquake waveform detection modules (green line). (b) Confusion matrix elements for phase picking evaluation.

Not limited to (6), for $40 \leq \alpha_{\text{res}} < 50$, we simply fixed labeled P at the fifth second. Although some of the earthquake waveforms are not complete, it is possible that the deep learning algorithms could characterize them with the available background noise. A model input template is shown in Fig. 7, showing the target functions for the seismic phase picking and earthquake waveform detection module, and the confusion matrix elements for seismic phase picking performance evaluation.

APPENDIX C

FORMULATIONS OF MARCHING MWA

Let τ be the length of the marching window. First, we slice a base three-component earthquake waveform as a function of time, $\omega_{\text{base}}(\mathbf{t})$, with a length of $(\tau_{\text{backward}} + 60 + \tau_{\text{forward}})$ seconds. Other earthquake waveforms are then randomly superimposed after the earthquake waveforms on the base waveform orderly, within the range of $\mathbf{t} \in [\tau_{\text{backward}}, \tau_{\text{backward}} + 60]$, forming a triplet set of 60-s-long waveform after marching backwardly and forwardly [Fig. 1(c.1)]

$$\omega_{\text{backward}} = \omega_{\text{base}}(\mathbf{t}) \Big|_{-\tau_{\text{backward}}}^{-\tau_{\text{backward}}+60} \quad (7)$$

$$\omega_{\text{center}} = \omega_{\text{base}}(\mathbf{t}) \Big|_{\tau_{\text{backward}}}^{\tau_{\text{backward}}+60} \quad (8)$$

$$\omega_{\text{forward}} = \omega_{\text{base}}(\mathbf{t}) \Big|_{\tau_{\text{backward}}+60}^{\tau_{\text{backward}}+60+\tau_{\text{forward}}} \quad (9)$$

where $\mathbf{t} \in [-\tau_{\text{backward}}, \dots, 0, \dots, 60 + \tau_{\text{forward}}]$. In ω_{center} , each earthquake waveform is paired with P and S arrivals, while ω_{backward} and ω_{forward} might include time-clipped earthquake waveforms with unpaired P and S arrivals. Note that the SNR of the superimposed P arrivals must be larger than 1.5 on the vertical channel high-passed at 2 Hz, and the SNR of “mosaic joints,” where other earthquake waveforms start to superimpose, must be lower than 2 across all the channels on waveform high-passed at 2 Hz. The SNR limitations of P arrivals ensure their visibility, and those of mosaic joints prevent mosaic waveforms from having unnatural abrupt changes every superimposition.

ACKNOWLEDGMENT

The authors thank the Southern California Earthquake Data Center (SCEDC) for providing seismic recordings for this study. They also thank the National Center for High-performance Computing (NCHC), Taiwan, for providing computational and storage resources.

REFERENCES

- [1] H. Kanamori, "Real-time seismology and earthquake damage mitigation," *Annu. Rev. Earth Planet. Sci.*, vol. 33, no. 1, pp. 195–214, 2005, doi: [10.1146/annurev.earth.33.092203.122626](https://doi.org/10.1146/annurev.earth.33.092203.122626).
- [2] T. Katayama and S. Horiuchi, *Early Earthquake Warning (EEW) System: Overview*. Berlin, Germany: Springer, 2015, pp. 621–629, doi: [10.1007/978-3-642-35344-4_66](https://doi.org/10.1007/978-3-642-35344-4_66).
- [3] R. M. Allen and D. Melgar, "Earthquake early warning: Advances, scientific challenges, and societal needs," *Annu. Rev. Earth Planet. Sci.*, vol. 47, no. 1, pp. 361–388, 2019, doi: [10.1146/annurev-earth-053018-060457](https://doi.org/10.1146/annurev-earth-053018-060457).
- [4] R. M. Allen and H. Kanamori, "The potential for earthquake early warning in southern California," *Science*, vol. 300, no. 5620, pp. 786–789, May 2003.
- [5] H. Serdar Kuyuk, R. M. Allen, H. Brown, M. Hellweg, I. Henson, and D. Neuhauser, "Designing a network-based earthquake early warning algorithm for California: ElarmS-2," *Bull. Seismol. Soc. Amer.*, vol. 104, no. 1, pp. 162–173, Dec. 2013, doi: [10.1785/0120130146](https://doi.org/10.1785/0120130146).
- [6] A. Chung, I. Henson, and R. Allen, "Optimizing earthquake early warning performance: ElarmS-3," *Seismol. Res. Lett.*, vol. 90, pp. 727–743, Mar. 2019.
- [7] N.-C. Hsiao, Y. Wu, T.-C. Shin, L. Zhao, and T.-L. Teng, "Development of earthquake early warning system in Taiwan," *Geophys. Res. Lett.*, vol. 36, no. 5, 2009, doi: [10.1029/2008GL036596](https://doi.org/10.1029/2008GL036596).
- [8] D.-Y. Chen, N.-C. Hsiao, and Y.-M. Wu, "The Earthworm based earthquake alarm reporting system in Taiwan," *Bull. Seismol. Soc. Amer.*, vol. 105, no. 2A, pp. 568–579, 2015, doi: [10.1785/0120140147](https://doi.org/10.1785/0120140147).
- [9] Y. Wu and L. Zhao, "Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning," *Geophys. Res. Lett.*, vol. 331, no. 6, 2006, doi: [10.1029/2006GL026871](https://doi.org/10.1029/2006GL026871).
- [10] Y. Wu, H.-Y. Yen, L. Zhao, B.-S. Huang, and W.-T. Liang, "Magnitude determination using initial P waves: A single-station approach," *Geophys. Res. Lett.*, vol. 330, no. 5, 2006, doi: [10.1029/2005GL025395](https://doi.org/10.1029/2005GL025395).
- [11] M. Böse, E. Hauksson, K. Solanki, H. Kanamori, Y. Wu, and T. Heaton, "A new trigger criterion for improved real-time performance of onsite earthquake early warning in southern California," *Bull. Seismol. Soc. Amer.*, vol. 99, no. 2A, pp. 897–905, 2009.
- [12] T.-Y. Hsu *et al.*, "Comparing the performance of the NEEWS earthquake early warning system against the CWB system during the 6 February 2018 M_w 6.2 Hualien earthquake," *Geophys. Res. Lett.*, vol. 45, no. 12, pp. 6001–6007, 2018.
- [13] R. Allen, "Automatic phase pickers: Their present use and future prospects," *Bull. Seismol. Soc. Amer.*, vol. 72, no. 6B, pp. S225–S242, 1982.
- [14] M. Baer and U. Kradolfer, "An automatic phase picker for local and teleseismic events," *Bull. Seismol. Soc. Amer.*, vol. 77, no. 4, pp. 1437–1445, 1987.
- [15] R. Sleeman and T. van Eck, "Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings," *Phys. Earth Planet. Interiors*, vol. 113, nos. 1–4, pp. 265–275, Jun. 1999.
- [16] C. D. Saragiotis, L. J. Hadjileontiadis, and S. M. Panas, "PAI-S/K: A robust automatic seismic P phase arrival identification scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1395–1404, Jun. 2002.
- [17] S. Nippres, A. Rietbrock, and A. Heath, "Optimized automatic pickers: Application to the ANCORP data set," *Geophys. J. Int.*, vol. 181, no. 2, pp. 911–925, 2010.
- [18] Z. E. Ross and Y. Ben-Zion, "Automatic picking of direct P, S seismic phases and fault zone head waves," *Geophys. J. Int.*, vol. 199, no. 1, pp. 368–381, Oct. 2014.
- [19] A. Lomax, C. Satriano, and M. Vassallo, "Automatic picker developments and optimization: FilterPicker—A robust, broadband picker for real-time seismic monitoring and earthquake early warning," *Seismol. Res. Lett.*, vol. 83, no. 3, pp. 531–540, May 2012. [Online]. Available: <https://pubs.geoscienceworld.org/srl/article/83/3/531-540/143936>
- [20] A. Jurkevics, "Polarization analysis of three-component array data," *Bull. Seismol. Soc. Amer.*, vol. 78, no. 5, pp. 1725–1743, 1988.
- [21] A. Cichowicz, "An automatic S-phase picker," *Bull. Seismol. Soc. Amer.*, vol. 83, no. 1, pp. 180–189, 1993.
- [22] C. Baillard, W. C. Crawford, V. Ballu, C. Hibert, and A. Mangeny, "An automatic kurtosis-based P- and S-phase picker designed for local seismic networks," *Bull. Seismol. Soc. Amer.*, vol. 104, no. 1, pp. 394–409, 2014.
- [23] F. Hirose, K. Miyaoka, N. Hayashimoto, T. Yamazaki, and M. Nakamura, "Outline of the 2011 off the Pacific coast of Tohoku earthquake (M_w 9.0)—Seismicity: Foreshocks, mainshock, aftershocks, and induced activity," *Earth, Planets, Space*, vol. 63, pp. 513–518, Jul. 2011.
- [24] E.-J. Lee, W.-Y. Liao, D. Mu, W. Wang, and P. Chen, "GPU-accelerated automatic microseismic monitoring algorithm (GAMMA) and its application to the 2019 Ridgecrest earthquake sequence," *Seismol. Res. Lett.*, vol. 91, no. 4, pp. 2062–2074, Mar. 2020.
- [25] E.-J. Lee, D. Mu, W. Wang, and P. Chen, "Weighted template-matching algorithm (WTMA) for improved foreshock detection of the 2019 Ridgecrest earthquake sequence," *Bull. Seismol. Soc. Amer.*, vol. 110, no. 4, pp. 1832–1844, 2020.
- [26] Z. E. Ross, M.-A. Meier, and E. Hauksson, "P wave arrival picking and first-motion polarity determination with deep learning," *J. Geophys. Res.*, vol. 123, pp. 5120–5129, Jun. 2018.
- [27] S. Hara, Y. Fukahata, and Y. Iio, "P-wave first-motion polarity determination of waveform data in western Japan using deep learning," *Earth, Planets Space*, vol. 71, p. 127, Nov. 2019.
- [28] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, 2018, Art. no. e1700578.
- [29] Z. E. Ross, M.-A. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learning," *Bull. Seismol. Soc. Amer.*, vol. 108, no. 5A, pp. 2894–2901, 2018.
- [30] L. Zhu *et al.*, "Deep learning for seismic phase detection and picking in the aftershock zone of 2008 M_w 7.9 Wenchuan earthquake," *Phys. Earth Planet. Interiors*, vol. 293, Aug. 2019, Art. no. 106261.
- [31] W. Zhu and G. C. Beroza, "PhaseNet: A deep-neural-network-based seismic arrival-time picking method," *Geophys. J. Int.*, vol. 216, no. 1, pp. 261–273, 2019.
- [32] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection," *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, 2019.
- [33] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Commun.*, vol. 11, no. 1, p. 3952, 2020. [Online]. Available: <http://www.nature.com/articles/s41467-020-17591-w>
- [34] W. Liao, E. Lee, D. Mu, P. Chen, and R. Rau, "ARRU phase picker: Attention recurrent-residual U-Net for picking seismic P-and S-phase arrivals," *Seismol. Res. Lett.*, vol. 92, no. 4, pp. 2410–2428, Mar. 2021, doi: [10.1785/0220200382](https://doi.org/10.1785/0220200382).
- [35] L. Gulia and S. Wiemer, "Real-time discrimination of earthquake foreshocks and aftershocks," *Nature*, vol. 574, pp. 193–199, Oct. 2019.
- [36] L. Gulia, S. Wiemer, and G. Vannucci, "Pseudoprospective evaluation of the foreshock traffic-light system in Ridgecrest and implications for aftershock hazard assessment," *Seismol. Res. Lett.*, vol. 91, no. 5, pp. 2828–2842, 2020, doi: [10.1785/0220190307](https://doi.org/10.1785/0220190307).
- [37] W.-Y. Liao, E.-J. Lee, D. Mu, and P. Chen, "Toward fully autonomous seismic networks: Backprojecting deep learning-based phase time functions for earthquake monitoring on continuous recordings," *Seismol. Res. Lett.*, vol. 93, no. 3, pp. 1880–1894, May 2022, doi: [10.1785/0220210274](https://doi.org/10.1785/0220210274).
- [38] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.
- [39] M. Crawshaw, "Multi-task learning with deep neural networks: A survey," 2020, *arXiv:2009.09796*.
- [40] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.
- [41] C. Doersch and A. Zisserman, "Multi-task self-supervised visual learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2070–2079, doi: [10.1109/ICCV.2017.226](https://doi.org/10.1109/ICCV.2017.226).
- [42] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," 2019, *arXiv:1901.11504*.

- [43] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 1871–1880.
- [44] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.
- [45] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3367–3375.
- [46] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks," 2017, *arXiv:1711.02257*.
- [47] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "Stanford earthquake dataset (STEAD): A global data set of seismic signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019.
- [48] Academia Sinica, Institute of Earth Sciences. (1994). *Broadband Array in Taiwan for Seismology*. [Online]. Available: <https://www.fdsn.org/networks/detail/TW/>
- [49] D. Mu, E.-J. Lee, and P. Chen, "Rapid earthquake detection through GPU-based template matching," *Comput. Geosci.*, vol. 109, pp. 305–314, Dec. 2006.
- [50] Y. M. Wu *et al.*, "A high-density seismic network for earthquake early warning in Taiwan based on low cost sensors," *Seismol. Res. Lett.*, vol. 84, no. 6, pp. 1048–1054, 2013, doi: [10.1785/0220130085](https://doi.org/10.1785/0220130085).
- [51] K.-S. Liu, T.-C. Shin, and Y.-B. Tsai, "A free-field strong motion network in Taiwan: TSMIP," *Terr., Atmos. Ocean. Sci.*, vol. 10, no. 2, pp. 377–396, Nov. 1999.
- [52] E.-J. Lee, P. Chen, T. Jordan, P. Maechling, M. Denolle, and G. Beroza, "Full-3D tomography (F3DT) for crustal structure in southern California based on the scattering-integral (SI) and the adjoint-wavefield (AW) methods," *J. Geophys. Res., Solid Earth*, vol. 119, Aug. 2014.
- [53] A. D'Alessandro, S. Scudero, and G. Vitale, "A review of the capacitive MEMS for seismology," *Sensors*, vol. 19, no. 14, p. 3093, Jul. 2019.
- [54] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal augmentation to improve generalization of deep neural networks," in *Advances in Geophysics*, vol. 61. Amsterdam, The Netherlands: Elsevier, 2020, pp. 151–177. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0065268720300030>



Wu-Yu Liao received the bachelor's and master's degrees in Earth sciences from the National Cheng Kung University, Tainan, Taiwan, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Department of Earth Sciences.

His research interests include deep learning applications in seismology, especially combining deep learning and physics-based algorithms and real-world applications of deep learning products.



En-Jui Lee received the master's degree in Earth sciences from the State University of New York (SUNY) at Binghamton, Binghamton, NY, USA, in 2008, and the Ph.D. degree in geophysics from the University of Wyoming, Laramie, WY, USA, in 2013.

He is currently an Associate Professor with the Department of Earth Sciences, National Cheng Kung University, Tainan, Taiwan. His research interests focus on earthquake monitoring, seismic hazard analysis, full-wave seismic tomography, earthquake source inversions, and applications of machine learning in seismology.



Da-Yi Chen received the Ph.D. degree in geosciences from the National Taiwan University, Taipei, Taiwan, in 2015.

He has been working at the Seismological Center, Central Weather Bureau, Taipei, since 2006. He is currently a Section Chief with the Seismological Center and an Adjunct Assistant Professor with the Department of the Earth and Life Sciences, University of Taipei, Taipei. His research interests are in the areas of earthquake monitoring and earthquake early warning.



2008 and was promoted

Po Chen received the bachelor's degree in geophysics from Peking University, Beijing, China, in 2000, and the Ph.D. degree in geological sciences (seismology) from the University of Southern California, Los Angeles, CA, USA, in 2005.

He worked as a Post-Doctoral Research Associate at the Lamont-Doherty Earth Observatory, Columbia University, New York City, NY, USA, from 2006 to 2008. He joined the Department of Geology and Geophysics, University of Wyoming, Laramie, WY, USA, as an Assistant Professor in 2008 and was promoted to an Associate Professor in 2014.



Dawei Mu received the master's degree in geology from Capital Normal University, Beijing, China, in 2006, and the Ph.D. degree in geophysics from the University of Wyoming, Laramie, WY, USA, in 2015.

He was worked as a Post-Doctoral Researcher with the San Diego Supercomputing Center, San Diego, CA, USA. His research interests include large-scale earthquake simulation and high-performance computing.



Yih-Min Wu received the Ph.D. degree in geophysics from the National Central University, Taoyuan, Taiwan, in 1999.

In the past few decades, he has established as one of the leading figures in earthquake early warning. After working for Central Weather Bureau, Taipei, Taiwan, as a Research Fellow, he accepted a Faculty position at the National Taiwan University (NTU), Taipei, in 2004. He was promoted to a Professor in 2010 and a Distinguished Professor in 2012.

He has been a Jointed Research Fellow of the Institute of Earth Sciences, Academia Sinica, since 2017. He was invited to participate in the Europe Framework Program (FP6 & FP7) projects to develop the earthquake early warning systems for European cities (2006–2014). Nowadays, he tries to share his knowledge and assist other countries (India, Indonesia, China, South Korea, Vietnam, Nepal, and Bhutan) to build up their earthquake early warning system using low-cost sensors. He works in the fields of real-time seismology and seismotectonics.

Dr. Wu has been a member of the Editorial Board of *Earth, Planets and Space* (EPS) since 2015, *Scientific Reports* since 2019, *Sensors* since 2021, and *Vietnam Journal of Earth Sciences* since 2015.