Natural Hazards
https://doi.org/10.1007/s11069-024-06464-w

ORIGINAL PAPER

Estimation of ground motion parameters via multi-task deep neural networks

Fanchun Meng[1] · Tao Ren[1] · Enming Guo[2] · Hongfeng Chen[3] · Xinliang Liu[1] · Haodong Zhang[1] · Jiang Li[1]

Received: 31 January 2023 / Accepted: 29 January 2024

Abstract

Ground motion parameters are crucial characteristics in earthquake warning and earthquake engineering practice. However, the existing methods are time-consuming and labor-intensive. In this study, a multi-task approach (GMP-MT) based on a hard parameter sharing strategy and single station data is proposed to improve the overall estimation accuracy by jointly optimizing the estimation of peak ground acceleration (PGA) and peak ground velocity (PGV). In addition, this study reshapes the mean squared error by adjusting the weight of the loss according to the data distribution to solve the data imbalance. The developed network structure extracts not only the seismic features from various dimensions but also the spatial–temporal correlations from large-dimensional seismic data. The designed model is trained and tested based on the global three-component seismic waveform data recorded in the STanford EArthquake Dataset. Experimental results show that the correlation coefficients of PGA and PGV are above 90%, and the average errors are less than 0.19. The model has good stability, specifically insensitive to epicenter distance, hypocentral depth, and signal-to-noise ratio. Furthermore, the superiority of the model in terms of learning and fitting is demonstrated by comparison with several state-of-the-art models in the existing literature.

Keywords Massive tectonic earthquakes · Ground motion parameter · Multi-task · Attention

✉ Tao Ren
chinarentao@163.com

[1] Software College, Northeastern University, No. 195, Chuangxin Road, Hunnan District, Shenyang 110169, China

[2] Computer Science and Engineering College, Northeastern University, No. 195, Chuangxin Road, Hunnan District, Shenyang 110169, China

[3] China Earthquake Networks Center, No. 56 Sanlihe Road, Xicheng District, Beijing 100045, China

Published online: 04 March 2024

1 Introduction

Ground motion parameters, the most critical information in seismic hazard analysis, are typically divided into time and response domains. PGA and PGV, two core parameters in the time domain, are applied to calculate the intensity of ground and buildings affected during an earthquake (Derakhshani and Foruzan 2019). Thus, the rapid and accurate estimation of PGA and PGV earns emergency response time for the management of railroads, oil pipelines, gas pipelines, etc., and provides a reference for constructing subsequent earthquake projects (Otake et al. 2020).

In contemporary seismology, the calculation of PGA and PGV relies on peaks in the complete waveforms of an event. But the traditional methods for extracting peaks are time-consuming and susceptible to noise, which affects the timeliness of earthquake warning and rapid reporting. In addition, the poor generalization of the traditional calculation methods leads to the absence of a uniform method for calculating ground motion parameters in various countries and regions (Güllü and Erçelebi 2007; Luco and Cornell 2007), which causes inconvenience to the follow-up work. Therefore, a method for estimating ground motion parameters with a broader range of applicability needs to be discovered.

With the increase in earthquake data volume and computational power, deep learning (DL) is widely applied (Zhang et al. 2018; Wang et al. 2019; Hacıefendioğlu et al. 2021). Convolutional neural network (CNN) has been proven to be extremely effective in extracting earthquake phase features while demonstrating superior performance on data containing strong background noise (Zhu and Beroza 2019; Mousavi et al. 2020; Xiao et al. 2021), earthquake early warning (Lin et al. 2021; Münchmeyer et al. 2021; Wibowo et al. 2021; Zhang et al. 2021), earthquake magnitude regression (Lomax et al. 2019; Kuang et al. 2021), localization (Kriegerowski et al. 2019; Shen and Shen 2021), and classification (Saad et al. 2020; Liu et al. 2021). Furthermore, Mangalathu and Burton (2019) used long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) to achieve the classification of building complex damage caused by earthquakes. Instead of traditional seismic data, textual descriptions of damage are used, which provides a new insight for network input. However, gradient vanishing is a critical problem in LSTM when dealing with long-term seismic data.

The application of DL to ground motion parameters estimation can be divided into two primary categories. (1) A computational model of ground motion parameters is constructed based on a priori knowledge, and then, the key parameters in the model are estimated using the DL algorithm. Alavi and Gandomi (2011) applied ANN model and simulated annealing (SA) algorithm to derive a new generalized attenuation relationship for the ground motion parameters. ANN/SA requires setting the influencing factors of the ground motion parameters manually, which results in the performance being determined by definition/selection of the factors and inability to be generalized. (2) Waveforms or features are provided as input for direct estimation of ground motion parameters based on the DL algorithm. Derakhshani and Foruzan (2019) implemented the identification of ground motion parameters such as PGA, PGV, etc., based on DNN, and manually selected seismic features as input. However, the algorithm lacks the design for the seismic characteristics, and it is difficult to extract deep features with fewer model parameter. Jozinović et al. (2020) proposed an intensity-predicting method of ground motion via CNN, which weakens the practical value with the high reliance on multiple station data. In addition, the absence of temporal features leads to limitations in the

performance. To the best of the authors' knowledge, no research has achieved the estimation of multiple ground motion parameters simultaneously, relying only on a single station and waveform data.

A DL network (GMP-MT) is proposed in this study for the rapid estimation of PGA and PGV. GMP-MT addresses part of the problems that prevail in existing studies, such as low generalization and the inability to maintain superior performance on unbalanced data. The contributions of this study in particular are as follows:

1. GMP-MT adopts the hard parameter sharing strategy that exploits the correlations among tasks. The strategy enhances the transferability of deep features, thereby improving the overall performance and generalization of GMP-MT.
2. GMP-MT has an efficient extraction structure comprising pyramidal convolution and self-attention mechanism. This structure excels in extracting spatio-temporal features across multiple dimensions and scales, and effectively reducing the impact of noise on estimation.
3. To overcome this limitation of seismic data imbalance, a new loss function compensation algorithm is proposed in this study. The method sets weights based on the volume of data, which enables the network to focus less on labels with larger sample sizes.
4. The experimental results on STanford EArthquake Dataset (STEAD) (Mousavi et al. 2019) validate the efficacy. GMP-MT achieves simultaneous estimations of PGA and PGV with errors below 0.2, providing reliable support for its practical applications in the field of seismology.

1.1 Data

Three-component seismic waveform data from STEAD are adopted. The dataset includes velocity waveform data of 75,198 events recorded by 1302 stations, and the distribution of events is shown in Fig. 1. Since the obtained seismic records are velocity records, the

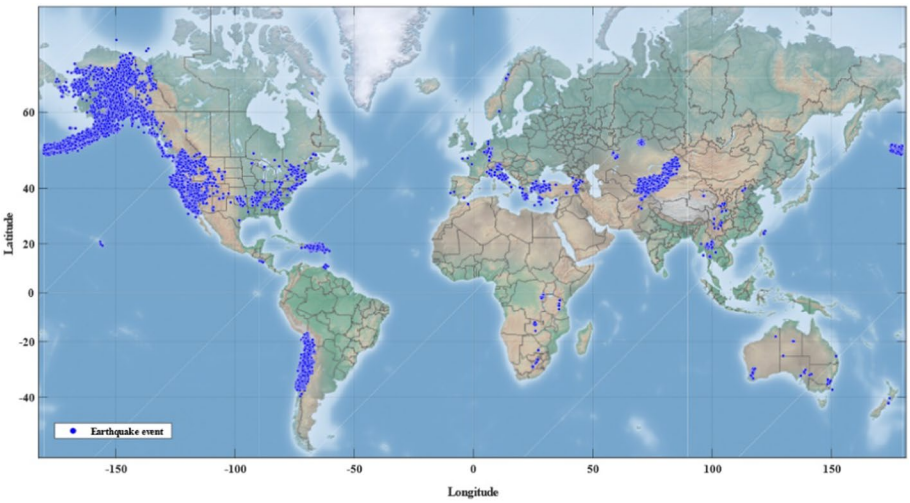


Fig. 1 Distribution of earthquake event. The event location is marked with a blue circular symbol

acceleration records are obtained by time derivatives. The synthesized three-directional velocity records and acceleration records are used separately:

$$v(t_i) = \sqrt{v^2_{\mathrm{E-W}}(t_i) + v^2_{\mathrm{N-S}}(t_i) + v^2_{\mathrm{U-D}}(t_i)}$$
$$a(t_i) = \sqrt{a^2_{\mathrm{E-W}}(t_i) + a^2_{\mathrm{N-S}}(t_i) + a^2_{\mathrm{U-D}}(t_i)}$$

where $v(t_i)$ and $a(t_i)$ are synthetic velocity records (m/s) and acceleration records (m/s^2) at time t_i; E–W, N–S, and U–D are east–west, north–south, and vertical directions. Then, PGV = max($v(t_i)$) and PGA = max($v(t_i)$). The $\log_{10}$(PGA) and $\log_{10}$(PGV) are adopted as training labels since the logarithmic function compresses the scale of the variables.

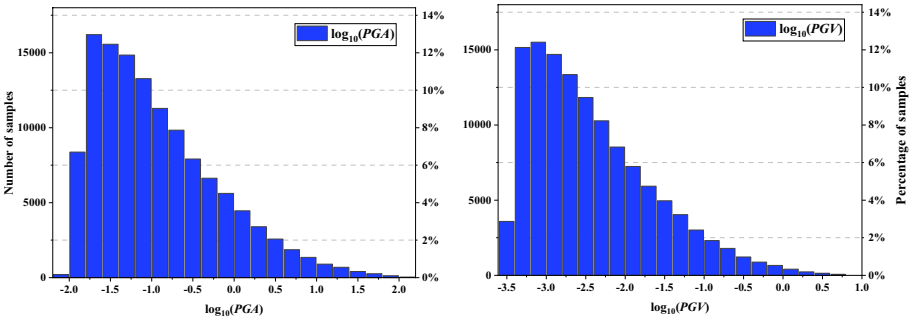
The hierarchical sampling based on the distribution of $\log_{10}$(PGA) divides the data into the training set, validation set, and test set with 131,300, 36,619, and 18,310, respectively. The distribution of $\log_{10}$(PGA) and $\log_{10}$(PGV) is shown in Fig. 2a and b. The selected waveform includes various types of seismic records from different regions worldwide, covering a wide range of epicentral distances, hypocentral depths, and signal-to-noise ratio (SNR). The distribution of epicenter distance, hypocentral depth, and SNR is shown in Fig. 2c–e. The diverse data provide a comprehensive performance assessment for this research and make the rapid estimation of PGA and PGV more challenging.

In this study, 10-s waveform data are adopted. The waveform is 1 s before and 9 s after the P-wave arrive time, as shown in Fig. 3. Since the number of samples is a critical factor affecting the performance of the DL, and it is difficult to determine in time whether the waveforms already contain the maximum value in an application, this paper retained the records (about 6%) that already contained the maximum value. As the interception method of the BHE and BHN channels is the same as that of the BHZ channel, only the interception of the BHZ channel is shown. Since the sampling frequency of the data is 100 Hz, a matrix of [3, 1000] is formed as the input of the GMP-MT.

1.2 GMP-MT model structure

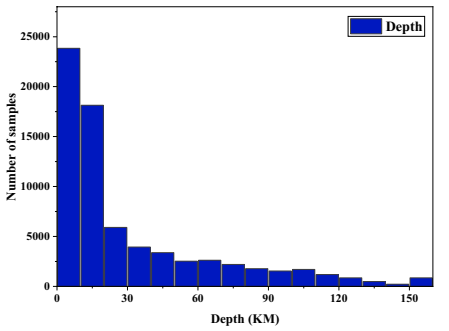
The GMP-MT model is a multi-task structure consisting of a deep encoder and two independent decoders. The hard parameter sharing strategy (Wibowo et al. 2021) is adopted to design the multi-task structure since the similarity between the two estimation tasks allows the similar optimization direction of the parameters in share layers. Shared layers are adopted as feature extractors to obtain common seismic features, while task-specific layers are adopted as feature resolvers to estimate PGA or PGV through seismic features.

In order to effectively extract the features embedded in the seismic data, this subsection proposes a pyramidal convolutional structure with nine convolutional layers, which adjusts the size of the convolutional kernels with different abstraction levels of the data to effectively extract features and reduces the computational load of large-dimensional raw seismic data. Each structure layer comprises a convolutional layer, a Leaky ReLU, and a pooling layer. Batch normalization (BN) (Ioffe and Szegedy 2015) is added in the 6th and 7th layers of the network. Residual block (He et al. 2016) is adopted in the 8th and 9th layers of the network to prevent gradient vanishing and overfitting. In addition, the high sampling frequency of current seismometers results in high dimensionality of seismic data within a short duration. The algorithms that solve temporal input, such as LSTM, may suffer from gradient vanishing when handling long-term data. An attention mechanism is incorporated to compute the interaction among features (Vaswani

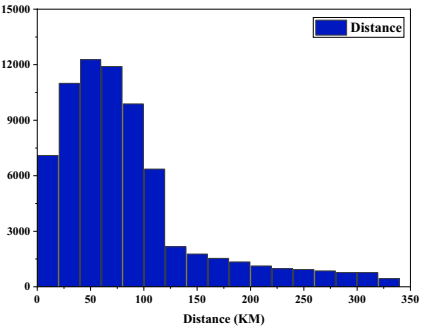


(a) Distribution of $\log_{10}(PGA)$

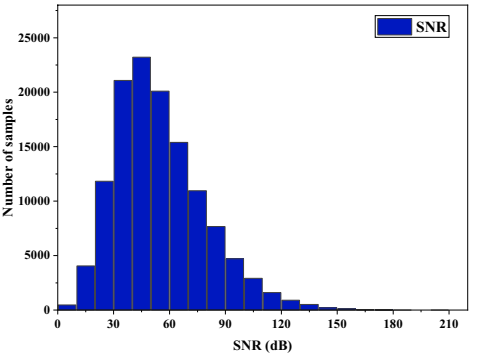
(b) Distribution of $\log_{10}(PGV)$



(c) Distribution of hypocentral depth



(d) Distribution of epicenter distance



(e) Distribution of SNR

Fig. 2 Data distribution

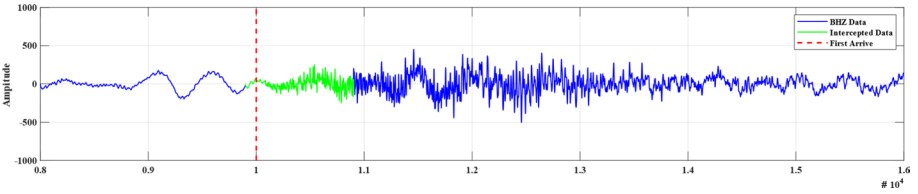


Fig. 3 Waveform data interception

et al. 2017), which overcomes the vanishing gradient, thus maintaining the features in the temporal dimension. The computation process of the attention mechanism is shown in Fig. 4. Specifically, the input size is $64 \times 1 \times 242$, where 64, 1, and 242 represent the number of channels, height, and width, respectively. Therefore, there are 242 points in the spatial domain. One-dimensional convolution is applied to process the input and obtain three tensors: q, k, and v. The similarity between each pair of points is obtained by computing the dot product of q and k^{T}. Then, a softmax operation is applied to produce the weight matrix of attention. The weights are multiplied with v to obtain tensor z, where tensor z is processed by the fully connected layer, which is used to increase the nonlinearity of the network. Ultimately, tensor z is stitched with the seismic feature map to obtain the final output tensor. The residual connection contributes to improving the flow of gradients during the backward pass.

The input to GMP-MT is a 3×1000 matrix, where the rows correspond to the 1000 samples from the BHE, BHN, and BHZ components of the seismic data. The outputs are $\log_{10}$(PGA) and $\log_{10}$(PGV). The open-source machine learning library PyTorch is adopted to establish and train the model. Figure 5 shows the complete structure of the GMP-MT model.

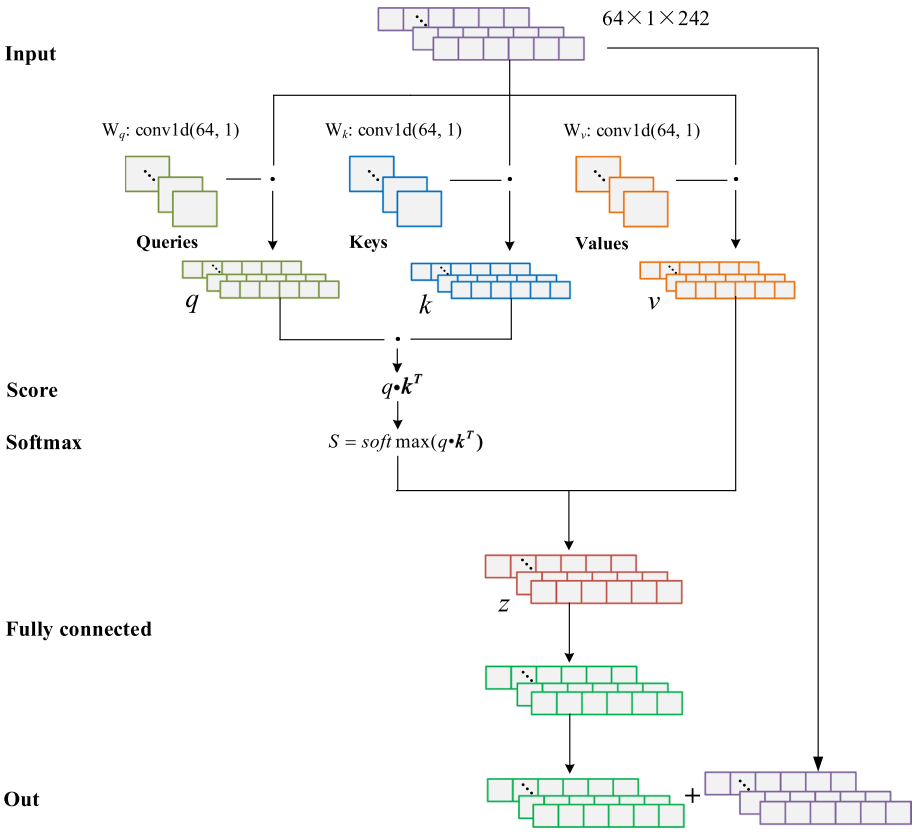


Fig. 4 Calculation of self-attention

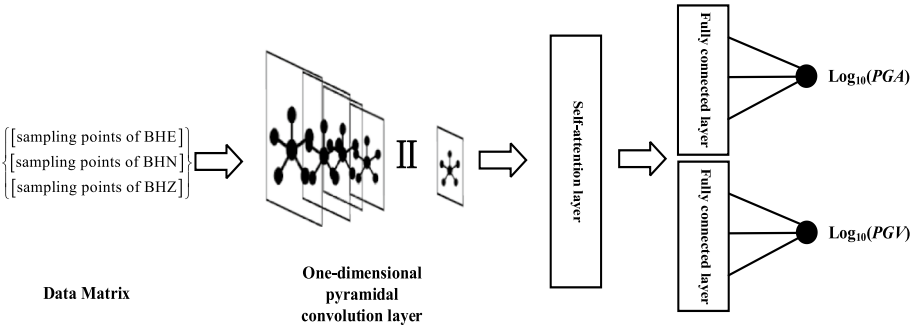


Fig. 5 Structure of GMP-MT

1.3 The loss function of the GMP-MT model

Statistically, most seismic events are low-intensity events. Thus, the extremely unbalanced distribution of the labels leads to a wrong optimization direction that focuses more on labels with larger sample sizes. This study proposes a new loss function, an improvement of MSELoss, to balance the samples by assigning higher weights to low-proportion labels and vice versa. This improvement can effectively reduce the search space of the model.

Firstly, the weights a_{PGA} and a_{PGV} are calculated for each range of data according to the distribution of $\log_{10}$(PGA) and $\log_{10}$(PGV), as shown in Eqs. (1) and (2). The $\mathrm{Num}_{\log10(\mathrm{PGA})}$ and $\mathrm{Num}_{\log10(\mathrm{PGV})}$ denote the number of $\log_{10}$(PGA) and $\log_{10}$(PGV), respectively. In order to focus the network on samples with smaller numbers, the body of the calculation adopts the inverse of the sample size. As the amount of data in different categories varies greatly, the logarithmic function is used to limit weights to avoid a huge gap in weights. In addition, the design of the weights ensures that the gradient is not zero during the parameter update process, so the denominator used $\log_{10}(1+\mathrm{Num}_{\log10(\mathrm{PGA})})$ instead of $\log_{10}(\mathrm{Num}_{\log10(\mathrm{PGA})})$. In this study, the training data are divided into eight sets, and the weights corresponding to each set are shown in Table 1.

$$a_{\mathrm{PGA}} = \frac{1}{\log_{10}(1+\mathrm{Num}_{\log_{10}(\mathrm{PGA})})} \tag{1}$$

$$a_{\mathrm{PGV}} = \frac{1}{\log_{10}(1+\mathrm{Num}_{\log_{10}(\mathrm{PGV})})} \tag{2}$$

The calculated weights are combined with MSELoss, as shown in Eqs. (3)–(8). The improved loss function assigns the weights more reasonably according to the amount of data, thus correcting the wrong optimization direction caused by data imbalance.

Table 1 Table of data weights

	[−4, −3]	[−3, −2]	[−2, −1]	[−1, 0]	[0, 1]	[1, 2]	[2, 3]
$\log_{10}$(PGA)		0.17	0.16	0.17	0.18	0.22	0.29
$\log_{10}$(PGV)	0.24	0.23	0.24	0.29	0.41		

$$\mathrm{Loss}_{\mathrm{PGA}} = a_{\mathrm{PGA}} \times \mathrm{MSE}_{\mathrm{PGA}} + L1 \tag{3}$$

$$\mathrm{Loss}_{\mathrm{PGV}} = a_{\mathrm{PGV}} \times \mathrm{MSE}_{\mathrm{PGV}} + L1 \tag{4}$$

$$\mathrm{MSE}_{\mathrm{PGA}} = \frac{1}{n}\sum (o_{\mathrm{PGA}}(x) - t_{\mathrm{PGA}}(x))^2 \tag{5}$$

$$\mathrm{MSE}_{\mathrm{PGV}} = \frac{1}{n}\sum (o_{\mathrm{PGV}}(x) - t_{\mathrm{PGV}}(x))^2 \tag{6}$$

$$L1 = \lambda \sum \omega_j^2 \tag{7}$$

$$\mathrm{Loss} = \mathrm{Loss}_{\mathrm{PGA}} + \mathrm{Loss}_{\mathrm{PGV}} \tag{8}$$

In Eqs. (3)–(8), $\mathrm{MSE}_{\mathrm{PGA}}$ and $\mathrm{MSE}_{\mathrm{PGV}}$ denote the MSELoss used to calculate $\log_{10}$(PGA) and $\log_{10}$(PGV), respectively; $L1$ denotes the lasso regularization; $\mathrm{o}_{\mathrm{PGA}}(x)$ and $\mathrm{o}_{\mathrm{PGV}}(x)$ denote the output of the network, respectively; $t_{\mathrm{PGA}}(x)$ and $t_{\mathrm{PGV}}(x)$ represent the true data labels, respectively; λ denotes the regularization parameter, which can be adjusted according to the training effect; and ω denotes the weights within the network.

2 Experimental analysis

Mean absolute error (MAE), mean squared error (MSE), and correlation coefficient (R^2) are chosen to evaluate the validity of the model. The evaluation metrics are calculated as Eqs. (9)–(11). The R^2 is an evaluation metric applied to evaluate the degree of fitting, with a result closer to 1, which indicates that the model is effective.

$$\mathrm{MAE} = \frac{1}{N}\sum\nolimits_{i=1}^{N} (|o_i - t_i|) \tag{9}$$

$$\mathrm{MSE} = \frac{1}{N}\sum\nolimits_{i=1}^{N} (o_i - t_i)^2 \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (t_i - o_i)^2}{\sum_{i=1}^{N} (t_i - \bar{t})^2} \tag{11}$$

In Eqs. (9)–(11), o_i, t_i, and $\bar{t}$ denote the output of the network model, the true target value corresponding to the data, and the mean of the true labels, respectively. Adam is used as the optimization function to train the model. The initial value of the learning rate is set to 10^{-3}. Considering the convergence of the model, a linear decay strategy is adopted for the learning rate to prevent oscillations and approach the optimum faster. The learning rate variation curves of the first 20 rounds are shown in Fig. 6.

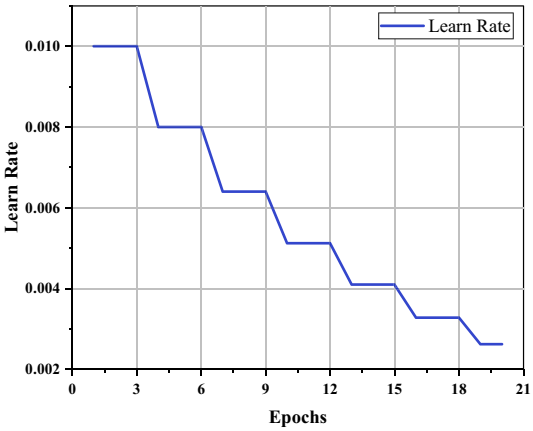


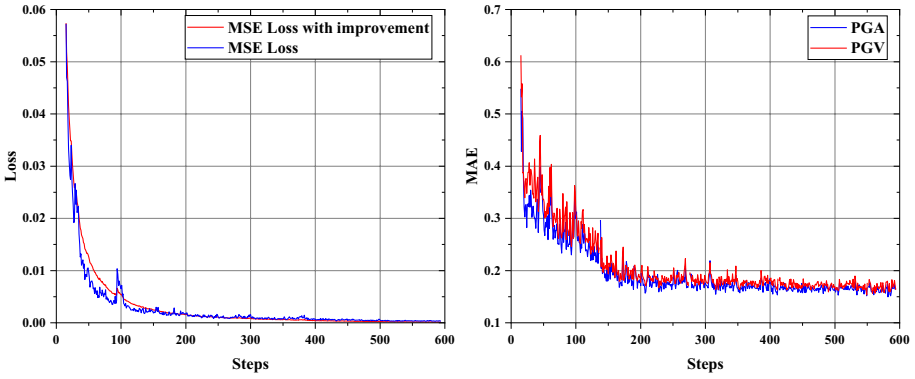
Fig. 6 Learning rate change curve

2.1 Model performance

The performance of GMP-MT is evaluated in this subsection. The training curves for the two loss functions, as well as the MAE variation curves obtained based on the improved loss function are shown in Fig. 7. Since the initial decline of the curve is rapid, making subsequent small changes difficult to discern, we start plotting the curve at the 7th training step.

The two curves in Fig. 7a are based on the same dataset, optimizer (Adam), and loss function decay strategy. The improvement of the loss function effectively mitigates the volatility of the training process and makes the training more stable. Meanwhile, the estimation performance of GMP-MT is improved. To highlight the applicability and anti-overfitting, the results of the model on the training and test sets are presented in Table 2, and the correlation coefficient of the model on the test set is shown in Fig. 8.

According to Table 2, GMP-MT excellently completes the estimation of PGA and PGV. It is worth noting that 94% of the waveform recordings used in the experiment do not contain the peaks of this event, which proves that besides the peaks, GMP-MT discovers



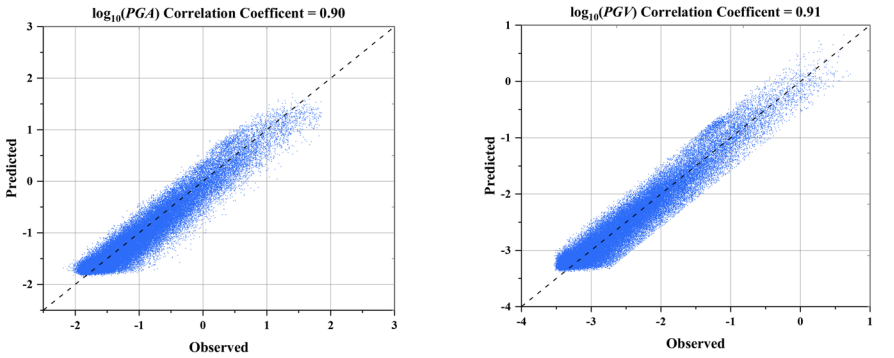
(a) Loss and Loss with improvement Trends over Training

(b) MAE Trends over Training Steps

Fig. 7 Training loss and MAE change curve

Table 2 The performance of the model

	Log_{10}(PGA)		Log_{10}(PGV)	
	Training	Testing	Training	Testing
MAE	0.16	0.17	0.18	0.19
MSE	$0.6*10^{-3}$	0.05	$0.6*10^{-3}$	0.05
R^2	0.96	0.90	0.96	0.91



(a) The Correlation coefficient of $\log_{10}(PGA)$ (b) The Correlation coefficient of $\log_{10}(PGV)$

Fig. 8 The correlation coefficient of each label

other features that help in the estimation of ground motion parameters. GMP-MT fits the nonlinear relationships contained in seismic waveforms, which are difficult to capture in traditional methods, by means of a multi-layer, multi-scale nonlinear transformation. The experiments achieve the similar MAE, MSE, and R^2, which proves that GMP-MT is able to extract hidden features shared by the two tasks. The excellent feature extraction is attributed to the fusion of deep and shallow features by pyramidal convolution. In contrast with the structure without fusion, pyramidal convolution prevents the feature space from being restricted to local features. In addition, on the two datasets, the training and test set, the MAE of GMP-MT only differs by 0.01. The design of 9-layer convolution not only avoids overfitting the model due to more parameters but also improves the nonlinear fitting ability of the model. We further note that the improved loss function suppresses the interference of data proportions to the model and intervenes in the focus of model learning by increasing the weights.

The distance of the epicenter and depth of the hypocenter can significantly affect the quality of seismic data due to the attenuation of seismic waves. Experiments are carried out to verify the sensitivity of GMP-MT to epicenter distance and hypocentral depth. The variation of the absolute error with the epicenter distance and hypocentral depth is provided in Fig. 9. It can be seen that a part of the data is selected by a stratified sampling method based on the data distribution.

Figure 9 indicates that about 85% of the absolute errors remain below 0.3, and about 65% stay below 0.2. It is particularly noteworthy that in the case of close-distance earthquakes (distance $\leq$ 200 km), the GMP-MT effectively maintains 95% of the errors below 0.3. The seismic signals undergo energy attenuation as the distance increases, and underground influencing factors such as material density make the attenuation patterns

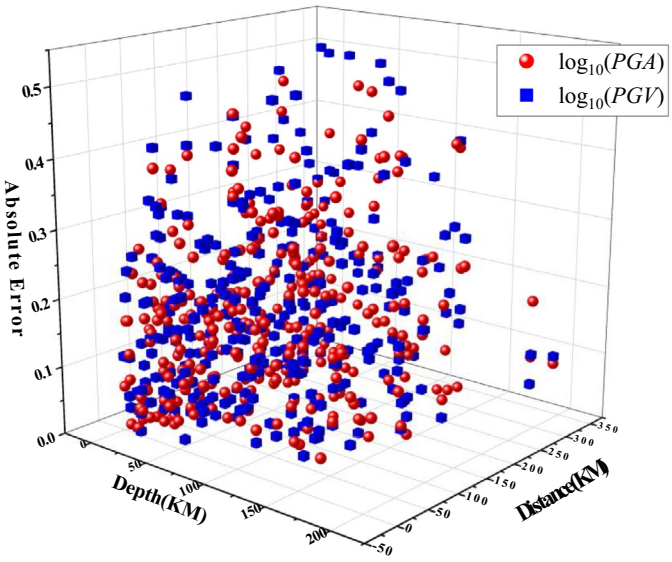


Fig. 9 Variation of errors with epicenter distance and hypocentral depth. The error distribution of $\log_{10}$(PGA) is marked by the sphere, and $\log_{10}$(PGV) is marked by the cube

become complex and diverse, resulting in the failure of GMP-MT to capture the complex energy attenuation patterns. Consequently, with the growing epicentral distance, the partial estimation error increases from 0.1 to 0.4, especially for the estimation of long-distance earthquakes (distance > 200 km). The multi-layer and multi-scale feature extraction structure enables GMP-MT to extract not only fundamental features but also more abstract higher-order features. Due to the distinctive structure, the model maintains the estimation error below 0.3 for deep-source earthquakes (depth > 100 km). For long-distance shallow earthquakes, the partial estimation error rises above 0.4. This can be attributed to the susceptibility of shallow-source earthquakes to geological interferences, resulting in signal variations during the propagation process surpass the fitting capabilities of DL.

Noise has a significant impact on the estimation task. However, the high sensitivity of seismic detectors makes waveforms susceptible to noise. The estimation model is required to have excellent anti-noise capability. To further verify the effect of noise on GMP-MT, the test set is divided into 19 groups according to the SNR distribution. The results are shown in Fig. 10.

The results in Fig. 10 indicate that the estimation error is less than 0.2 on low SNR data (SNR < 80 db), which demonstrates the outstanding anti-noise capability of GMP-MT. The pyramidal convolutional structure of GMP-MT is able to capture and separate the noise at multiple scales, such as high frequency, low frequency, and so on. At the same time, with the self-attention mechanism, it completes the feature extraction of the real signal, thereby diminishing the interference of noise. However, as the SNR increases and the signal quality improves, the estimation error rises significantly. Combined with Fig. 2e, this phenomenon is intricately linked to data distribution, and fewer high SNR data pose challenges for the model to learn extraction of key features from high SNR data. Hence, with the abundance of data, it is anticipated that the anti-noise property of GMP-MT will be further enhanced.

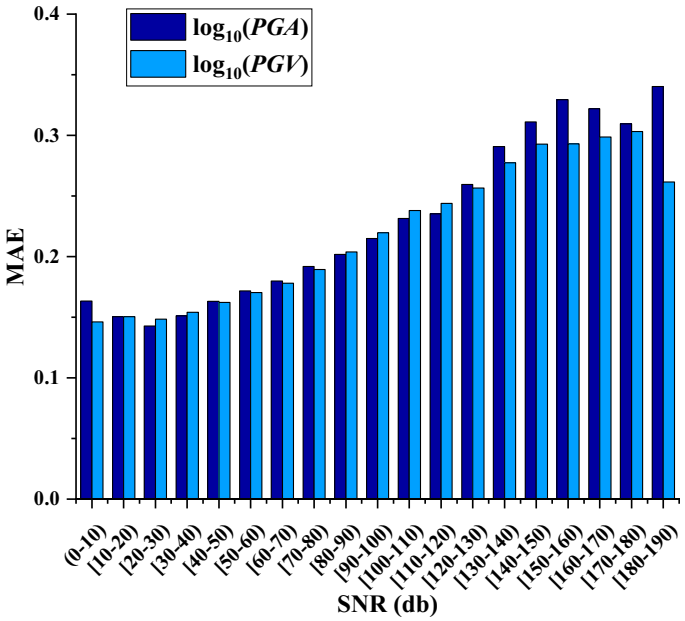


Fig. 10 Variation of errors with SNR

2.2 Ablation analysis

In this subsection, we verify whether the pyramidal convolutional and the self-attention are suitable for GMP-MT by testing in the ground motion parameter estimation task. We remove the pyramidal convolution and self-attention in GMP-MT, respectively. The comparison results of the various structures are shown in Table 3.

According to Table 3, self-attention contributes the most to the GMP-MT model, which proves that the temporal features of the waveform are the key to the identification of ground motion parameters. Self-attention is capable of picking up the relationship between seismic features and obtaining weights of different features indicating the importance each characteristic account for. In addition, when pyramidal convolution is removed from the model, a small performance degradation occurs due to poor spatial feature extraction. Waveform data are commonly large in dimensionality, and the information of individual points in

Table 3 Results of ablation experiments

Model structure		Evaluation metrics					
Pyramidal convolutional	Attention	Log_{10}(PGA)			Log_{10}(PGV)		
		MAE	MSE	R^2	MAE	MSE	R^2
✓		0.46	2.6	0.47	0.45	2.6	0.5
	✓	0.20	0.1	0.87	0.21	0.11	0.89
✓	✓	0.17	0.05	0.90	0.19	0.05	0.91

waveform is scarce; therefore, it is necessary to reduce the dimensionality of the features through pyramidal convolution.

2.3 Generalizability verification

In this subsection, the generalization of the GMP-MT model is verified. Considering that the complex and diverse topography of China can comprehensively evaluate the generalization of the model, a total of 49,790 seismic data from 2009 to 2017 within China provided by the China Earthquake Network Center (CENC) are used in this experiment. The experimental results of the GMP-MT model are shown in Table 4.

As shown in Table 4, the performance of GMP is still outstanding on the Chinese data, which can be attributed to the ability to extract part of deep general features of seismic data in pyramidal convolution and attention mechanism. The GMP-MT model attempts to extract generic features in waveform data for ground motion parameter estimation, thus providing strong constraints on the possible decision boundaries for model learning (Zhu et al. 2020), and this constraint is the key element for the accurate estimation of ground motion parameters. However, due to the diversity of this dataset, the errors in the Chinese data increase slightly.

3 Discussion

3.1 Results of comparative experiments

In this subsection, the superiority of the proposed algorithm is highlighted by comparing it with the state-of-the-art algorithms listed in Table 5. All algorithms are tested based on the STEAD public dataset and run on the hardware environment of Intel Core i5-13600KF, 32 GB RAM, and NVIDIA GeForce RTX 3060Ti 8 GB. The term "N/A" indicates that the corresponding value is unavailable.

Based on the statistical metrics listed in Table 5, the combination of pyramidal convolutional and self-attention makes GMP-MT superior to existing state-of-the-art algorithms in terms of learning and fitting. The GP/OLS, MEP, and GP/SA are based on genetic programming to discover the mapping relationships between ground motion parameters and magnitude, shear wave velocity, etc. However, the relationships are limited to specific regions and cannot be generalized to the global STEAD dataset. Furthermore, they exhibit suboptimal performance for low-magnitude events ($M_w < 4$). The ANN/SA and DNN are incorporated with artificial neural networks (ANN), but are limited by the number of parameters. The ANN fails to extract the features effectively, leading to a deterioration in estimation accuracy when faced with more complex seismic source mechanisms. The CNN algorithms adopt DL, which maintains outstanding performance in the more diverse

Table 4 Result of generalizability experiments

Model	Log_{10}(PGA)			Log_{10}(PGV)		
	MAE	MSE	R^2	MAE	MSE	R^2
GMP-MT (STAND)	0.17	0.05	0.90	0.19	0.05	0.91
GMP-MT (CHINA)	0.19	0.06	0.86	0.27	0.05	0.90

Table 5 Comparison of experimental results

	PGA			PGV		
	R^2	MAE	RMSE	R^2	MAE	RMSE
ANN/SA (Alavi and Gandomi 2011)	N/A	1.03	1.16	N/A	2.38	2.49
GP/OLS (Gandomi et al. 2011)	N/A	1.71	2.1	N/A	0.73	0.89
MEP (Alavi et al. 2011)	N/A	1.23	1.50	N/A	1.85	2.09
GP/SA (Kafaei Mohammadnejad et al. 2012)	N/A	1.75	1.97	N/A	1.41	1.63
DNN (Derakhshani and Foruzan 2019)	**0.91**	0.40	0.50	0.89	0.44	0.56
CNN (Jozinović et al. 2020)	0.90	0.37	**0.24**	0.89	0.36	0.23
GMP-MT	0.90	**0.17**	0.25	**0.91**	**0.19**	**0.22**

Bold values indicate the optimal values for each metric

STEAD dataset, but it loses important temporal features compared to GMP-MT. The combined structure in GMP-MT improves the nonlinear fitting ability and enables the model to mine deep spatiotemporal features. In addition, the correlation coefficient values indicate that the improved loss function effectively prevents overfitting. GMP-MT is also proven to span a subspace of high-dimensional space under limited samples, so that the model completes a low-error regression on individual anomalous data.

3.2 Analysis of hard parameter sharing

In multi-task networks, feature sharing plays a key role among related tasks, enabling the network to better extract generalized features. Consequently, compared to multiple independent single-task networks, multi-task networks are more efficient. In general, multi-task networks contain two different parameter sharing strategies, hard parameter sharing, and soft parameter sharing, as depicted in Fig. 11. In this study, the hard parameter sharing strategy is applied to both the PGA and PGV estimation tasks, which means that they share the feature extraction component while retaining the task-specific feature parsing layers. Conversely, the soft parameter sharing strategy (Ma et al. 2018) involves each task having independent feature extraction layers with regularization to ensure similarity among parameters of these layers.

This subsection validates the effect of the hard parameter sharing multi-task learning strategy. The experiments are conducted around four models: the GMP-MT model, the single-task GMP-MT-PGA model for estimating PGA, the GMP-MT-PGV model for

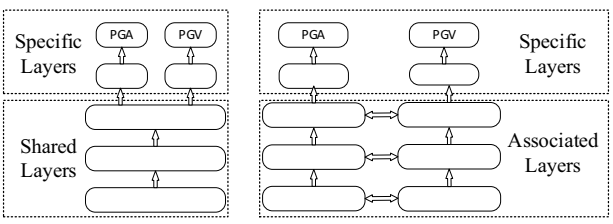


Fig. 11 Parameter sharing strategy

estimating PGV, and the GMP-MT-soft model with soft parameter sharing strategy. Among them, the GMP-MT-X model for single task is consistent with the GMP-MT model except for the task-specific layer. The number of parameters in the model and running time for each estimation are recorded to evaluate the performance of the models. The experimental results are summarized in Table 6.

The results show that the GMP-MT model with the multi-tasking framework achieves the lowest error. The multi-task framework takes full advantage of the fact that PGA and PGV have similar optimization directions by combining different tasks to optimize the feature extractor of GMP-MT so that the extracted features can accomplish estimations of deeper features. In addition, these two tasks limit each other and prevent overfitting to one task. Compared with the single-task model, although GMP-MT adds an extra task feature resolution layer, the number of parameters and the running time are smaller than the sum of the two single-task models. It can be seen that the GMP-MT applying the soft parameter sharing strategy has a 70% increase in network parameters and a 58% increase in inference time due to the addition of the feature extraction component. In addition to MAE, GMP-MT based on hard parameter sharing shows better performance. In contrast with MAE, the deterioration of MSE and R2 demonstrates that GMP-MT based on soft parameter sharing is not adept at dealing with anomalous data in the dataset, i.e., poor generalization. In soft parameter sharing, different tasks have private feature extraction components. Although the parameters are maintained similar through strategies, each feature extraction component is still biased to extract features that are more favorable to the current task. Hard parameter sharing compensates for this problem by co-training the same feature extraction layer with losses from different tasks, so that the extracted features are more generalized and generic.

4 Conclusions

This study addresses the challenges associated with ground motion parameter estimation methods, which include dependence on maximum magnitude, poor generalization, and susceptibility to noise. In this study, we introduce the GMP-MT, a multi-task deep learning framework, as a solution for the simultaneous and fast estimation of PGA and PGV. Our findings emphasize the significance of spatio-temporal feature fusion, especially in terms of model generalizability and anti-interference. The combined structure of pyramidal convolution and self-attention mechanism not only extracts multi-scale spatial features but also solves the problem of gradient vanishing which exists in time-domain feature extraction. The experimental results show that GMP-MT is able to leverage the fact that PGA

Table 6 Result of multi-task experiments

Model	Evaluation metrics							
	Params	Time (s)	Log_{10}(PGA)			Log_{10}(PGV)		
			MAE	MSE	R^2	MAE	MSE	R^2
GMP-MT-PGA	0.22 M	7.66×10^{-5}	0.20	0.07	0.85	**N/A**	**N/A**	**N/A**
GMP-MT-PGV	0.22 M	7.66×10^{-5}	**N/A**	**N/A**	**N/A**	0.19	0.06	0.88
GMP-MT (Soft)	0.38 M	1.48×10^{-4}	**0.16**	0.13	0.87	0.21	0.14	0.83
GMP-MT	0.24 M	9.37×10^{-5}	0.17	**0.05**	**0.90**	**0.19**	**0.05**	**0.91**

and PGV have similar optimization directions to effectively reduce the time and number of parameters required for estimating ground motion parameters.

This research underscores the potential of multi-task deep learning as a valuable tool in ground motion parameter estimation, with implications for improving the efficiency and accuracy of seismic data analysis. The proposed model is being prepared for deployment at CENC as one of the auxiliary tools for rapid earthquake early warning and rapid reporting. In addition, the possibility of using shorter data could be explored.

Acknowledgements This work is partially supported by National Natural Science Foundation of China (62276058, 61902057, and 41774063), Fundamental Research Funds for the Central Universities (N2217003), and Joint Fund of Science and Technology Department of Liaoning Province and State Key Laboratory of Robotics, China (2020-KF-12-11).

Author contributions Fanchun Meng helped in formal analysis, methodology, software, validation, visualization, and writing—original draft. Tao Ren helped in conceptualization, funding acquisition, resources, supervision, writing—review, and editing. Enming Guo helped in methodology. Hongfeng Chen helped in data curation and resources. Xinliang Liu helped in methodology and supervision. Haodong Zhang contributed to writing—review and editing. Jiang Li helped in methodology and editing.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability Waveform data, metadata, or data products for STEAD can be downloaded from https://github.com/smousavi05/STEAD. The Chinese dataset is from CENC. We only have the right to use data from CENC and cannot share it with others. If the reader is interested in these data, he/she can contact CENC (zouly@seis.ac.cn) for permission.

Code availability The GMP-MT algorithm was developed based on Windows 10, pycharm, Python 3.6, and PyTorch 1.10. The algorithm size is 22 MB. The application may be downloaded from https://github.com/Fan-Chun-Meng/GMP-MT. For enquiries contact chinafcmeng@163.com.

Declarations

Competing interests The authors declare no conflict of interest related to this work.

References

Alavi AH, Gandomi AH (2011) Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. Comput Struct 89(23–24):2176–2194. https://doi.org/10.1016/j.compstruc.2011.08.019

Alavi AH, Gandomi AH, Modaresnezhad M, Mousavi M (2011) New ground-motion prediction equations using multi expression programing. J Earthq Eng 15(4):511–536. https://doi.org/10.1080/13632469.2010.526752

Derakhshani A, Foruzan AH (2019) Predicting the principal strong ground motion parameters: a deep learning approach. Appl Soft Comput 80:192–201. https://doi.org/10.1016/j.asoc.2019.03.029

Gandomi AH, Alavi AH, Mousavi M, Tabatabaei SM (2011) A hybrid computational approach to derive new ground-motion prediction equations. Eng Appl Artif Intell 24(4):717–732. https://doi.org/10.1016/j.engappai.2011.01.005

Güllü H, Erçelebi E (2007) A neural network approach for attenuation relationships: an application using strong ground motion data from Turkey. Eng Geol 93(3–4):65–81. https://doi.org/10.1016/j.enggeo.2007.05.004

Hacıefendioğlu K, Başağa HB, Demir G (2021) Automatic detection of earthquake-induced ground failure effects through Faster R-CNN deep learning-based object detection using satellite images. Nat Hazards 105:383–403. https://doi.org/10.1007/s11069-020-04315-y

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Francis B, David B (eds) Proceedings of the 32nd international conference on machine learning. Proceedings of machine learning research, PMLR, vol 37, pp 448–456

Jozinović D, Lomax A, Štajduhar I, Michelini A (2020) Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. Geophys J Int 222(2):1379–1389. https://doi.org/10.1093/gji/ggaa233

Kafaei Mohammadnejad A, Mousavi SM, Torabi M, Mousavi M, Alavi AH (2012) Robust attenuation relations for peak time-domain parameters of strong ground motions. Environ Earth Sci 67:53–70. https://doi.org/10.1007/s12665-011-1479-9

Kriegerowski M, Petersen GM, Vasyura-Bathke H, Ohrnberger M (2019) A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. Seismol Res Lett 90(2A):510–516. https://doi.org/10.1785/0220180320

Kuang W, Yuan C, Zhang J (2021) Network-based earthquake magnitude determination via deep learning. Seismol Res Lett 92(4):2245–2254. https://doi.org/10.1785/0220200317

Lin JT, Melgar D, Thomas A, Searcy J (2021) Early warning for great earthquakes from characterization of crustal deformation patterns with deep learning. J Geophys Res Solid Earth 126(10):e2021JB022703. https://doi.org/10.1029/2021JB022703

Liu X, Ren T, Chen H, Chen Y (2021) Classification of tectonic and non-tectonic seismicity based on convolutional neural network. Geophys J Int 224(1):191–198. https://doi.org/10.1093/gji/ggaa444

Lomax A, Michelini A, Jozinović D (2019) An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. Seismol Res Lett 90(2A):517–529. https://doi.org/10.1785/0220180311

Luco N, Cornell CA (2007) Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. Earthq Spectra 23(2):357–392. https://doi.org/10.1193/1.2723158

Ma J, Zhao Z, Yi X, Chen J, Hong L, Chi Ed H (2018) Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1930–1939. https://doi.org/10.1145/3219819.3220007

Mangalathu S, Burton HV (2019) Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions. Int J Disaster Risk Reduct 36:101111. https://doi.org/10.1016/j.ijdrr.2019.101111

Mousavi SM, Sheng Y, Zhu W, Beroza GC (2019) STanford EArthquake Dataset (STEAD): a global data set of seismic signals for AI. IEEE Access 7:179464–179476. https://doi.org/10.1109/ACCESS.2019.2947848

Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat Commun 11(1):3952. https://doi.org/10.1038/s41467-020-17591-w

Münchmeyer J, Bindi D, Leser U, Tilmann F (2021) The transformer earthquake alerting model: a new versatile approach to earthquake early warning. Geophys J Int 225(1):646–656. https://doi.org/10.1093/gji/ggaa609

Otake R, Kurima J, Goto H, Sawada S (2020) Deep learning model for spatial interpolation of real-time seismic intensity. Seismol Soc Am 91(6):3433–3443. https://doi.org/10.1785/0220200006

Saad OM, Hafez AG, Soliman MS (2020) Deep learning approach for earthquake parameters classification in earthquake early warning system. IEEE Geosci Remote Sens Lett 18(7):1293–1297. https://doi.org/10.1109/LGRS.2020.2998580

Shen H, Shen Y (2021) Array-based convolutional neural networks for automatic detection and 4D localization of earthquakes in Hawai ‘i. Seismol Res Lett 92(5):2961–2971. https://doi.org/10.1785/0220200419

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30

Wang B, Zhang N, Lu W, Wang J (2019) Deep-learning-based seismic data interpolation: a preliminary result. Geophysics 84(1):V11–V20. https://doi.org/10.1190/geo2017-0495.1

Wibowo A, Pratama C, Sahara DP, Heliani L, Rasyid S, Akbar Z, Muttaqy F, Sudrajat A (2021) Earthquake early warning system using ncheck and hard-shared orthogonal multitarget regression on deep learning. IEEE Geosci Remote Sens Lett 19:1–5. https://doi.org/10.1109/LGRS.2021.3066346

Xiao Z, Wang J, Liu C, Li J, Zhao L, Yao Z (2021) Siamese earthquake transformer: a pair-input deep-learning model for earthquake detection and phase picking on a seismic array. J Geophys Res Solid Earth 126(5):e2020JB021444. https://doi.org/10.1029/2020JB021444

Zhang G, Wang Z, Chen Y (2018) Deep learning for seismic lithology prediction. Geophys J Int 215(2):1368–1387. https://doi.org/10.1093/gji/ggy344

Zhang X, Zhang M, Tian X (2021) Real-time earthquake early warning with deep learning: application to the 2016 M 60 Central Apennines, Italy earthquake. Geophys Res Lett 48(5):2020GL089394. https://doi.org/10.1029/2020GL089394

Zhu W, Beroza GC (2019) PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophys J Int 216(1):261–273. https://doi.org/10.1093/gji/ggy423

Zhu W, Mousavi SM, Beroza GC (2020) Seismic signal augmentation to improve generalization of deep neural networks. Adv Geophys 61:151–177. https://doi.org/10.1016/bs.agph.2020.07.003