'DiTing' and 'DiTingtools':a large multi-label dataset and algorithm set for intelligent seismic data processing established based on the China Seismological Network

Ming Zhao[12], Zhuowei Xiao [3] , Shi Chen[12], Bo Zhang[12], Bei Zhang[12], Honglei Li[12]

1. Institute of Geophysics，China Earthquake Administration，Beijing 100081，China;
2. Beijing Baijiatuan Earth Science National Observation and Research Station，Beijing 100095，China;
3. Institute of Geology and Geophysics, Chinese Academy of Sciences,Beijing 100029,China

ABSTRACT

A set of software and hardware systems with artificial intelligence (AI for short) technology as the core has been developed, which includes: "DiTing" large-scale AI seismology training dataset, AI algorithms named "DiTingTools" for automatic seismic data processing, edge device named "DiTing" wisdom box to deploy AI algorithms at the edge and closely interacting with the cloud-based post-processing module. These work further promoted the application of AI in seismology as well as the big data real-time monitoring of China Seismological Network.

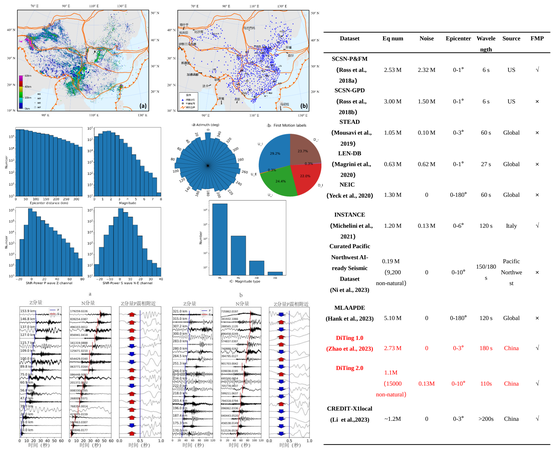
DATA AND METHOD

◆Background

- **In recent years, AI technology has developed rapidly in the field of seismic signal recognition, showing great potential and is expected to replace traditional seismic data processing methods. The develop and practical applications of AI are inseparable from massive, high-quality annotated data. However, the current work on AI data sets in the field of seismology is still in its infancy. In particular, until recently there is no such work in China. The China Seismological Network has currently built more than 1,300 high-standard fixed stations and over 15000 stations of all types, forming the world's largest monitoring and early warning network. Besides, China owns a large and professional earthquake cataloging team to guarantee the quality of manual cataloging and phase reports. These two points give China unique advantages in developing AI datasets and algorithms for seismic big data processing applications.**

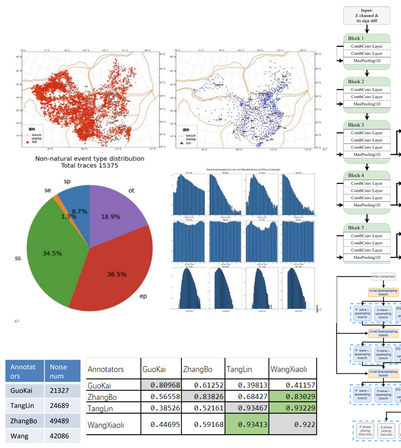
◆Data description

- **Based on the seismic bulletin from China Earthquake Network Center (CENC) and the waveform data from the International Earthquake Science Data Center(IESDC), we developed the DiTing dataset V1.0 & V2.0. The datasets contain approximately 3.83 million three-component seismic waveforms and corresponding P-wave and S-wave arrival tags, as well as corresponding earthquake type、noise type、phase type、 P-wave first motion polarity(FMP for short)、bac-azimuth、epicenter distance、magnitude tags(Zhao et al.,EQS2022).According to the data security policy, the dataset has been strictly desensitized, erasing information such as earthquake origin time and location, station name and location, to ensure that it can only be used to develop AI tools.**



Zhao M et al.,EQS2023

Dataset	Eq num	Noise	Epicenter	Wavelength	Source	FMP
SCSN-P&FM (Ross et al., 2018a)	2.53 M	2.32 M	0-1°	6 s	US	√
SCSN-GPD (Ross et al., 2018b)	3.00 M	1.50 M	0-1°	6 s	US	×
STEAD (Mousavi et al., 2019)	1.05 M	0.10 M	0-3°	60 s	Global	×
LEN-DB (Magrini et al., 2020)	0.63 M	0.62 M	0-1°	27 s	Global	×
NEIC (Yeck et al., 2020)	1.30 M	0	0-180°	60 s	Global	×
INSTANCE (Michelini et al., 2021)	1.20 M	0.13 M	0-6°	120 s	Italy	√
Curated Pacific Northwest AI-ready Seismic Dataset (Ni et al., 2023)	0.19 M (9,200 non-natural)	0	0-10°	150/180 s	Pacific Northwest	×
MLAAPDE (Hank et al., 2023)	5.10 M	0	0-180°	120 s	Global	×
DiTing 1.0 (Zhao et al., 2023)	2.73 M	0	0-3°	180 s	China	√
DiTing 2.0	1.1M (15000 non-natural)	0.13M	0-10°	110s	China	√
CREDIT-Xiliocal (Li et al.,2023)	~1.2M	0	0-3°	>200s	China	√



Zhao M et al.,FES2023

Annotators	Noise num
GuoKai	21327
TangLin	24689
ZhangBo	49489
Wang	42086

Annotators	GuoKai	ZhangBo	TangLin	WangXiaoli
GuoKai	0.80968	0.61252	0.39813	0.41157
ZhangBo	0.56558	0.83826	0.68427	0.83029
TangLin	0.38526	0.52161	0.93467	0.93229
WangXiaoli	0.44695	0.59168	0.93413	0.922

Fig.2 DiTing dataset v 2.0

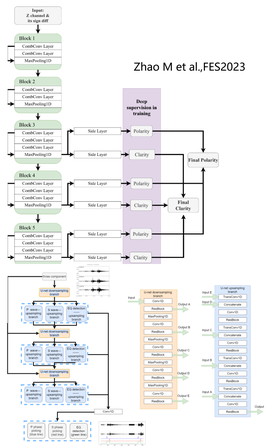
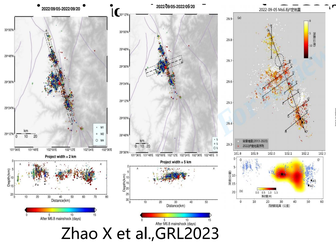


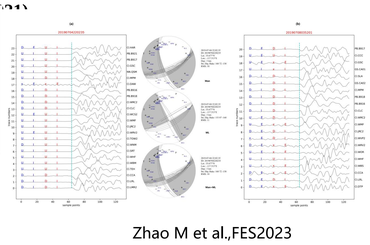
Fig.3 DiTingMotion and DiTingPicker

MAIN RESULTS

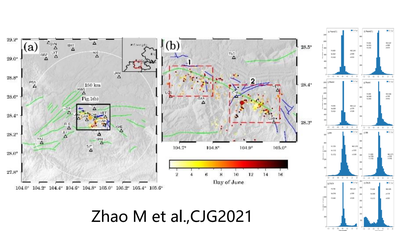
- **The release of DiTing dataset enhanced the generalization ability of the deep learning model on China Seismological Network data. With the support of DiTing dataset, we developed and trained several deep learning models referred as DiTingTools for automatic seismic data processing. The DiTingPicker has excellent generalization ability in P、S phase picking and is used in the automated location catalog of the Luding earthquake sequence(Zhao et al.,GRL2023).The DiTingMotion generalize well on P wave FMP identification both on domestic and international data.Thus we build a fully automatic focal mechanism inversion workflow for small and medium earthquakes—— the DiTing-FOCALFLOW(Zhao et al.,FES2023).The workflow proved to be efficient on the 2021 Yangbi earthquake aftershock sequence (ML≥2.0), and the 2019 Ridgecrest earthquake aftershock sequence (ML≥2.5).**
- **Because many AI models trained on international datasets have obvious decline in generalization ability on Chinese data, so we fine-tuned PhaseNet and applied it to obtain the high-resolution catalog that further expanded the detection limit of micro-seismic magnitude.We detected 4 to 5 times more small events than the routine catalog, which significantly enhanced microseismic activities monitoring capabilities in industrial mining areas.We applied the fine-tuned PhaseNet model to build foreshock catalog of the 2019 Ms 6.0 Changning earthquake,and revealed the abnormal microseismic activity before the Ms 6.0 mainshock area and better described the spatiotemporal evolution characteristics of**



Zhao X et al.,GRL2023



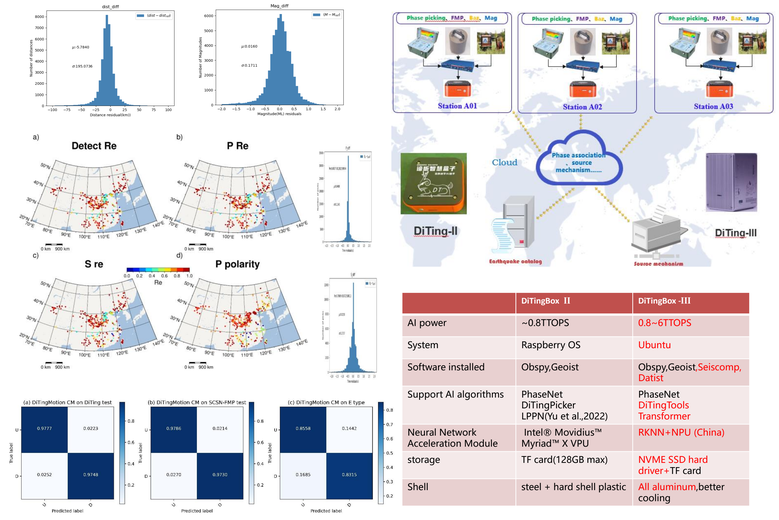
Zhao M et al.,FES2023



Zhao M et al.,CJG2021

APPLICATION

- **In the continuous waveform detection and evaluation of more than 1,000 stations over a year across China, DiTingPicker has achieved an average recall rate of 80% for event detection, mean square error ±0.2s for P phase picking, and ±0.4s for S, comparing with routnine catalog built by CENC . The DiTingMotion achieved average identification accuracy rate of Pg FMP reached 86.7% (U) and 87.9% (D), and 75.1% (U) and 73.1% (D) for Pn FMP, the average magnitude prediction error of a single station is mainly concentrated at ±0.5.**
- **We developed an edge computing device named DiTing Smart Box.Now the newest type is DiTing-III.It is small size, low power consumption, and able to achieve high-performance AI model inference at the edge. It is also compatible with popular seismological data processing software like obspy, seiscomp, and pretrained AI models like PhaseNet and DiTingTools. This allows us to process seismic data directly on the instrument side, and has broad prospects in earthquake early warning and real-time dynamic monitoring.**
- **With DiTing Smart Box and the cloud-based post-processing modules like phase association, earthquake location and relocation, focal mechanism inversion, a decentralized real-time seismic data system integrating edge and cloud was established,we deploy this system to the China Seismological Network with more than 1,000 stations, and intelligent foreshock monitoring is being carried out .**



	DiTingBox Ⅱ	DiTingBox -ⅢⅠ
AI power	~0.8TTOPS	0.8~6TTOPS
System	Raspberry OS	Ubuntu
Software installed	Obspy,Geoist	Obspy,Geoist,Seiscomp, Datist
Support AI algorithms	PhaseNet DiTingPicker LPPN(Yu et al.,2022)	PhaseNet DiTingTools Transformer
Neural Network Acceleration Module	Intel® Movidius™ Myriad™ X VPU	RKNN+NPU (China)
storage	TF card(128GB max)	NVME SSD hard driver+TF card
Shell	steel + hard shell plastic	All aluminum,better cooling

CONCLUSION

1. **We proposed China's first large-scale artificial intelligence seismology multi-task training dataset – "DiTing" .**
2. **We independently developed and trained a series of high-generalization AI algorithms: DiTingPicker to process earthquake detection and seismic phase picking which achieved high generalize ability on the CENC data; DiTingMotion to effectively identify P-wave first motion polarity, especially the recognition of "Emergent" FMP characteristics, has been significantly improved.**
3. **The DiTing-FOCALFLOW automatic focal mechanism inversion process is proposed, which can achieve quasi-real-time focal mechanism solution for small and medium earthquake. It can further be used for dynamic monitoring of aftershocks, identification of earthquake foreshock sequences, etc.**
4. **We developed an edge computing device that can deploy well-trained AI models for decentralize seismic data auto processing - the DiTing Smart Box.**
5. **We developed the DiTing edge-cloud integrated platform for real-time or offline seismic big data, which can obtain earthquake source parameter immediately after it occurs, providing support for the identification of earthquake precursor information and the rapid production of major earthquake emergency technology products.**

PAPERS AND ACHIEVEMENTS

• PAPERS：

- **Zhao M**, Xiao Z, Chen S, et al. DiTing: A large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology[J]. Earthquake Science, 2023, 36(2): 84-94.
- **ZHAO, M.**, TANG, L., Chen, S., SU, J., & ZHANG, M. (2021). Machine learning based automatic foreshock catalog building for the 2019 M S 6.0 Changning, Sichuan earthquake. Chinese Journal of Geophysics, 64(1), 54-66.
- **Zhao M**, Xiao Z, Zhang M, Yang Y, Tang L and Chen S (2023), DiTingMotion: A deep-learning first-motion-polarity classifier and its application to focal mechanism inversion. Front. Earth Sci. 11:1103914.doi: 10.3389/feart.2023.1103914.
- Zhao, X., Xiao, Z., Wang, W., Li, J., **Zhao, M.**, Chen, S., & Tang, L. (2023). Fine seismogenic fault structures and complex rupture characteristics of the 2022 M6. 8 Luding, Sichuan earthquake sequence revealed by deep learning and waveform modeling. Geophysical Research Letters, 50(18), e2023GL102976.

•DOI

- DiTing dataset 2.0： 10.11998/IESDC.SEIS.D03MOD20230002
- DiTing dataset 1.0: 10.12080/nedc.11.ds. 2022.0002 or CSTR:12166.11.ds.2022.0002
- **Ming Zhao**. (2020, November 4). mingzhaochina/unet_cea: unet_cea (Version 1.2). Zenodo. http://doi.org/10.5281/zenodo.4243864
- **Ming Zhao**, Jiahui Ma, Hao Chang. (2020, November 5). mingzhaochina/aicwps-pspicker 1.0 (Version 1.0). Zenodo. 10.5281/zenodo.4246086
- **Ming Zhao**. (2020, November 4). mingzhaochina/dataset: dataset for CNN and Unet (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.4243985