Seismic Phase Picking

Ruihuan Wang
School of Electronics Engineering and Computer Science
Peking University
dylanwr@stu.pku.edu.cn

Yuchen Wang
School of Electronics Engineering and Computer Science
Peking University
2100013153@stu.pku.edu.cn

Abstract

Seismic phase picking, which aims to determine the arrival time of P- and S-waves according to seismic waveforms, is fundamental to earthquake monitoring. Generally, manual phase picking is trustworthy, but with the increasing number of worldwide stations and seismic monitors, it becomes more challenging for human to complete the task comprehensively. In this work, we explore multiple ways to do automatic phase picking, including traditional and learning-based methods.

1 Introduction

Seismic phase picking and detection are the first major steps for earthquake localization and further seismology research. With an increasing number of seismic monitors and waveform data recorded by them, more efficient and robust tools are in urgent need.

For automatic phase picking, there are two mainstream types of method. One is traditional method[1, 2, 16], which apply well designed filters or matrix to the waveform arrays. Another is learning-based methods[13, 14, 19], which build machine learning models for the picking task. Network for multi-station phase picking[5, 8] also appear in recent years.

In this work, we try both traditional and learning-based models to do phase picking and analyze their performance. Due to time and computing resource constraint, out work mainly focuses on single-trace phase picking and some of our ideas are not tested, but we also propose them in Sec. 6.

2 Dataset

We use the INSTANCE dataset[12] for training and testing.

2.1 Basic information

The INSTANCE dataset consists of nearly 1.2 million 3-component waveform traces from about 50,000 earthquakes and more than 130,000 noise 3-component waveform traces. The waveform dataset is accompanied by metadata consisting of more than 100 parameters providing comprehensive information, and thus is suitable for machine learning analysis.

2.2 Features

In this work, we mainly focus on parameters `trace_P_arrival_sample` and `trace_S_arrival_sample`, which represent the arrival time of P- and S-waves of an

earthquake. Fig. 1 shows some distributional features of the dataset. In Sec. 4.1, we will further discuss how the features influence our data building work.
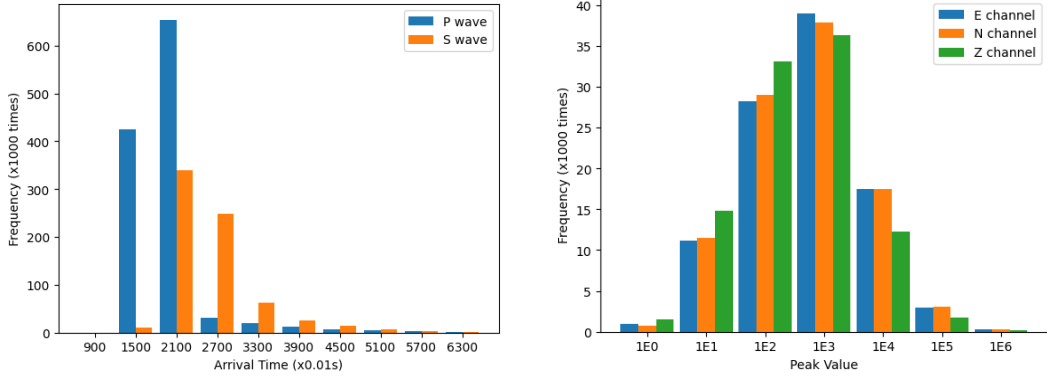
(a) Distribution of arrival time of P- and S-waves. Some of the waveform traces do not contain S-waves.

(b) Distribution of peak values in three channels (of 100,000 samples). Note that the value on x-axis is increasing exponentially.

Figure 1: A glance into distributional features of the INSTANCE dataset.

3 Traditional method

A classical traditional phase picking algorithm is AR-AIC model, which is widely used in related works (*e.g.* [1, 15]). We apply AR-pick tool from ObsPy[3] to 1,000 samples from INSTANCE and Tab. 1 shows our result.

Table 1: **Result of AR pick.** μ and σ are mean and standard deviation of errors (*i.e.* ground truth - prediction) in seconds respectively. Pr is precision. Arrival time residuals that are less than threshold are counted as true positive.

Phase	μ	σ	threshold (s)	Pr
			0.10	0.34
			0.30	0.63
P	0.86	8.84	0.50	0.71
			1.00	0.77
			2.00	0.80
			0.10	0.13
			0.30	0.28
S	-4.17	17.20	0.50	0.40
			1.00	0.64
			2.00	0.84

Generally, the performance of AR pick is not satisfying enough. (1) Fig. 1a shows that for most records, the arrival time are less than 40 seconds, but the mean of error can reach over 4 seconds for S-wave, which is an error of around 10%. (2) The standard deviation is relatively large, which demonstrates a lack of robustness. (3) Even when we set a threshold to 2s, Pr value still cannot reach 0.90.

Therefore, AR Pick is not capable of precise phase picking and cannot replace manual phase picking. A more powerful method is needed.

4 Learning-based method

4.1 Data preprocessing and label designing

The first task to do supervised learning is to build data-label pairs.

As shown in Fig. 2a, the data from INSTANCE have a uniform length of 12,000 samples (120 seconds with 100Hz sampling rate). But according to Fig. 1a, arrival time of signals are mainly distributed in

the range of 1,000 - 6,000. So it is reasonable to set a waveform window for training, which means to take the former 6,000 record samples and drop the rest. In this way we are able to accelerate the training progress while taking full advantage of the dataset.

Besides, from Fig. 1b, we see a relatively wide range of peak value distribution. So it is not wise to directly put raw data into the net. We normalize *each component* by

$$\text{Unbiased}_i = x_i - \overline{x_i},$$

$$\text{Normed}_i = \frac{\text{Unbiased}_i}{\max\left\{|\text{Unbiased}_i|\right\}}$$

where $i \in \{\text{E, N, Z}\}$ represents a certain channel and x_i is the raw value of the i-th component.

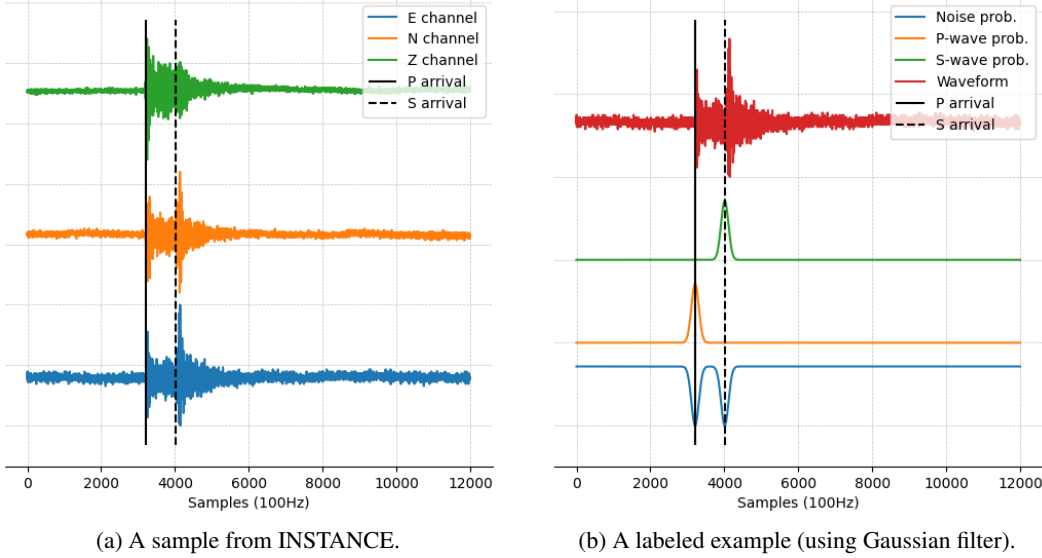| (a) A sample from INSTANCE. | (b) A labeled example (using Gaussian filter). |

Figure 2: An example of data and labeling.

As for labels, an intuitive idea is to treat the phase picking task as a regression problem, and label each waveform with 2 numbers (arrival time of P- and S- wave). But there are some challenges if using this method. (1) Not all traces contain both types of wave. Sometimes the station fail to record S-wave because S-wave cannot travel through the molten outer core of the Earth. Besides, when coping with pure noise waveform, such labeling seems impossible to be done. (2) The arrival time has uncertainty and it is improper to see the arrivals as single numbers, so we expect the ground truth will be centered on the manual picks with some uncertainty.

For this reason, we apply a mask which reach peak value at the manual picked time point and the nearby data points have reduced values (Fig. 2b shows an example labeled by Gaussian filters. We also try triangle and box filters in our experiment), representing the probability of arrival. Data points other than arrival of P- and S-waves are considered noise. Thus the probability distribution of noise can be calculated by

$$\text{Prob}_{\text{noise}} = 1 - \text{Prob}_{\text{P}} - \text{Prob}_{\text{S}},$$

where Prob means the probability of each class.

4.2 Models

4.2.1 CNN-based model: PhaseNet

The architecture of PhaseNet (Fig. 3) is modified from U-net[14], with 4 down-sampling and 4 up-sampling stages. 1-D convolutions and ReLU activations are applied inside each stage. A skip connection that directly concatenates the left output to the right layer is also used to improve convergence. In the last layer, the softmax function is used to calculate probabilities of noise, P-wave and S-wave.
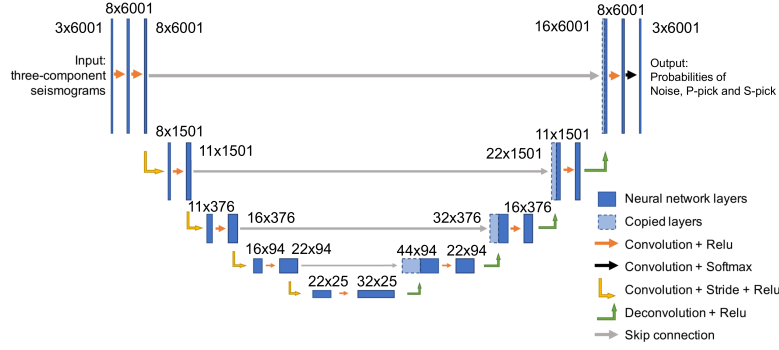
3

Figure 3: PhaseNet architecture (layer dimensions modified by us) from [19].

4.2.2 Attention-based model: EQTransformer

As shown in Fig. 4, EQTransformer consists of a deep encoder and three separate decoders that detect earthquake, pick P- and S-wave respectively.

4.3 Loss

The loss function is defined as cross-entropy between the ground truth distribution (y) and predicted distribution (y') to measure the divergence between them:

$$\mathcal{L}(y, y') = -\sum_{i=1}^{3} \sum_{\text{sample points}} y_i(x) \log(y_i'(x) + \varepsilon),$$

where $i = 1, 2, 3$ represents noise, P- and S-wave. ε is added to avoid taking the logarithm of 0 and is set to 1E-5 in our experiment.

5 Experiment

The experiment progress does not go as smooth as expected.

We first try to implement and train PhaseNet. Since it is originally designed to cope with waveform with a length of 3,000, we do a dimension modification first (see Fig. 3). However, it is strange that the model does not converge correctly. We manage to print out the predicted result and find that all values of probabilities of P- and S-waves are 0, which means the model seems to regard all sample points as noise. We have made lots of efforts to solve this, such as adding/removing layers (like Dropout and BatchNormalization), trying different strategies to preprocess the data, trying different sets of kernel sizes, increase the weights of wave layers when computing loss, etc. Yet eventually the problem remains unsolved.

There is also a possibility that there are bugs in our training progress due to our lack of deep learning practical experience on such a large dataset.

As for EQTransformer, we also implement a PyTorch version and manage to fit the pretrained model into our torch version implementation. Though not trained on INSTANCE, the result (see Tab. 2) still beat AR Pick on this dataset, which shows a high generalization and reliability of learning-based picker.
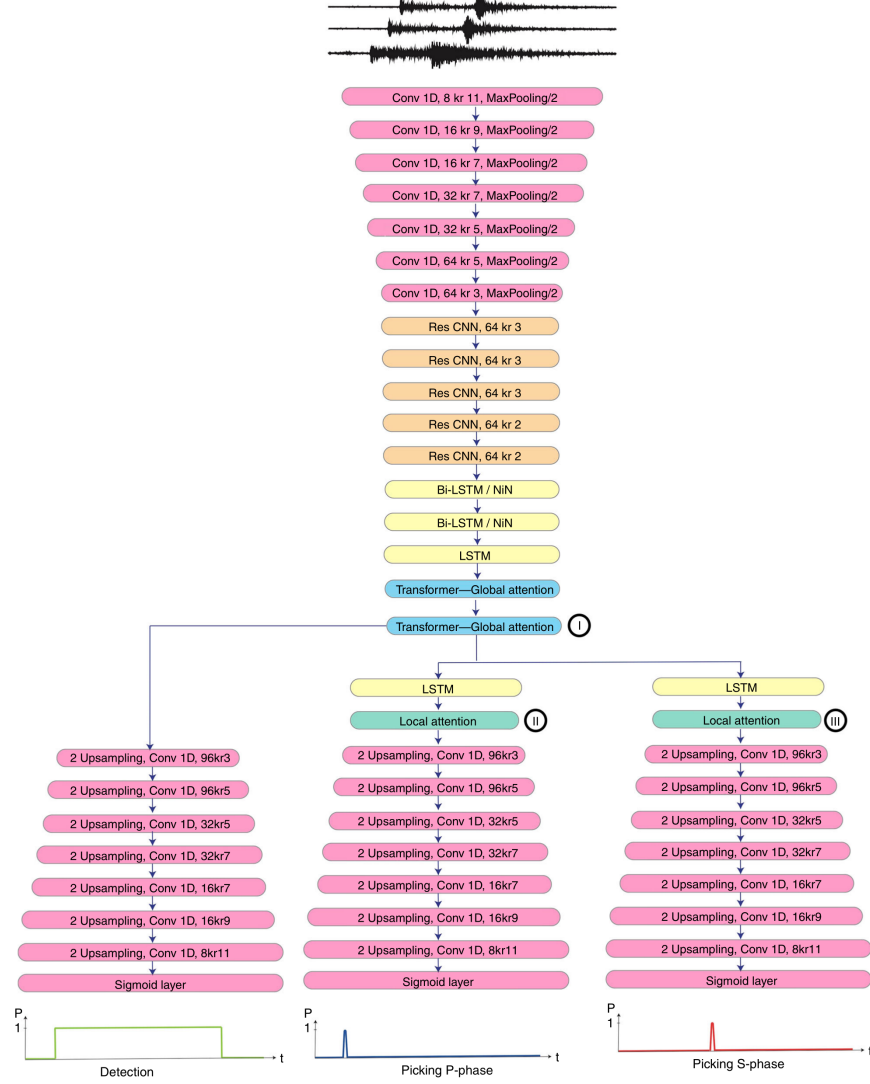
Figure 4: EQTransformer architecture (taken from [13]).

Table 2: **Comparison of EQTransformer and AR Pick.** μ and σ are mean and standard deviation of errors (*i.e.* ground truth - prediction) in seconds respectively. Pr, Re, F1 are precision, recall and F1-score. Arrival time residuals that are less than $\Delta t = 0.5$s are counted as true positive. When using AR Pick, we simply enforce it to predict the arrival time of both P- and S-wave, no matter whether there is corresponding wave. So Re and F1 of AR Pick are meaningless here.

Phase	Model	μ	σ	Pr	Re	F1
P	AR Pick	0.86	8.84	0.71	-	-
	EQTransformer	-0.27	4.18	0.76	1.00	0.86
S	AR Pick	-4.17	17.20	0.40	-	-
	EQTransformer	-0.39	5.37	0.61	0.88	0.72

Fig. 5 shows some visualized results of EQTransformer and AR Pick. In each figure, only the E channel of waveform is shown. We can clearly understand how noise influences the prediction. When there is little noise (Fig. 5a and Fig. 5b), both models work well. When noise gets stronger (Fig. 5c and Fig. 5d), AR Pick may fail to give a correct prediction. If the amplitude of noise is very close to the waves (Fig. 5e and Fig. 5f), both models cannot deal with the condition very well, and it is also challenging for humans to pick the phases out if only given single trace. This is the reason why
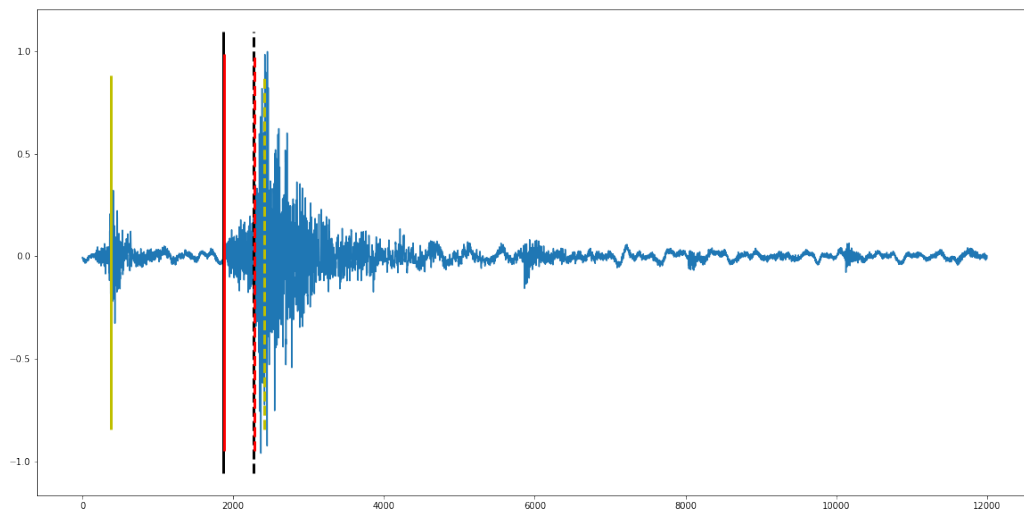
we need multi-trace based phase picking techniques. Through analyzing the waveforms of the same earthquake recorded by multiple stations, the models are able to get more robust to noise.
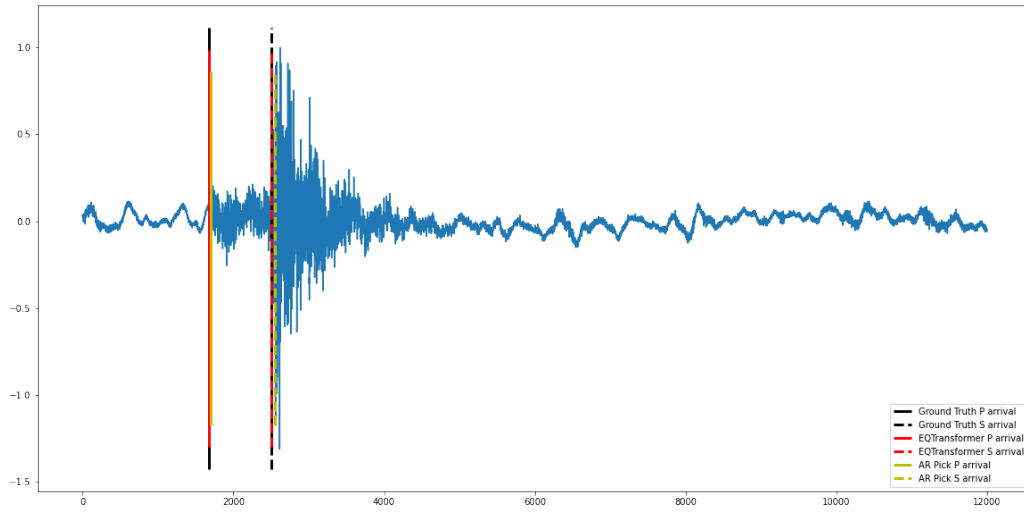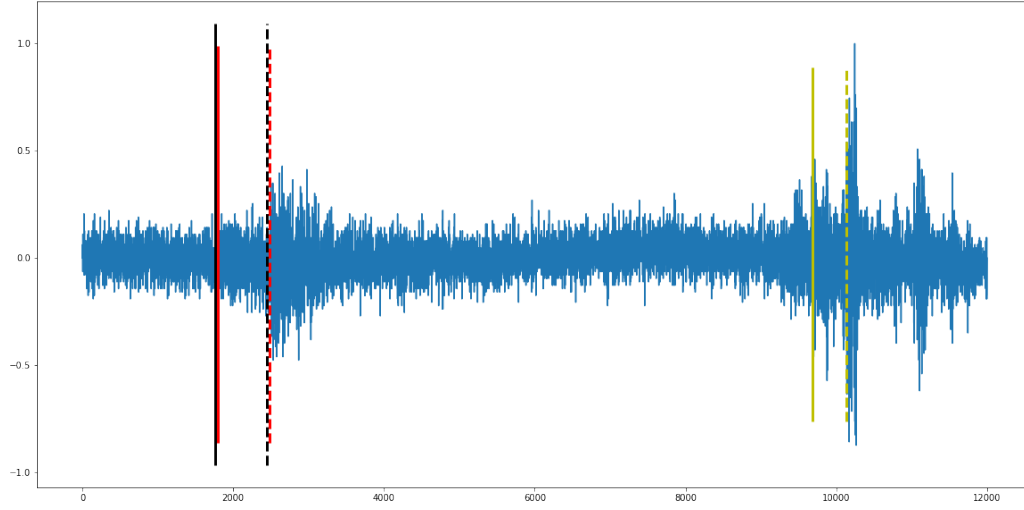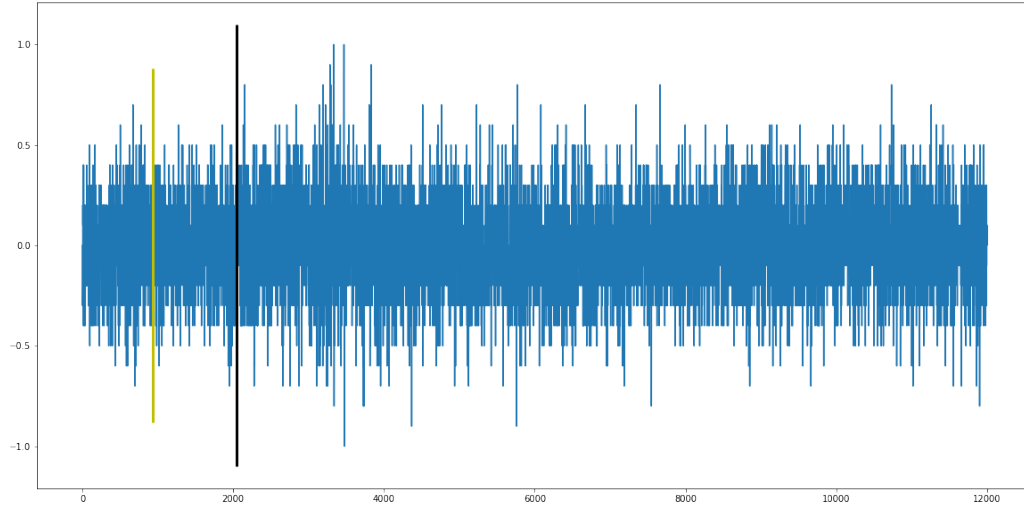

(a) Slight noise, both models work well.
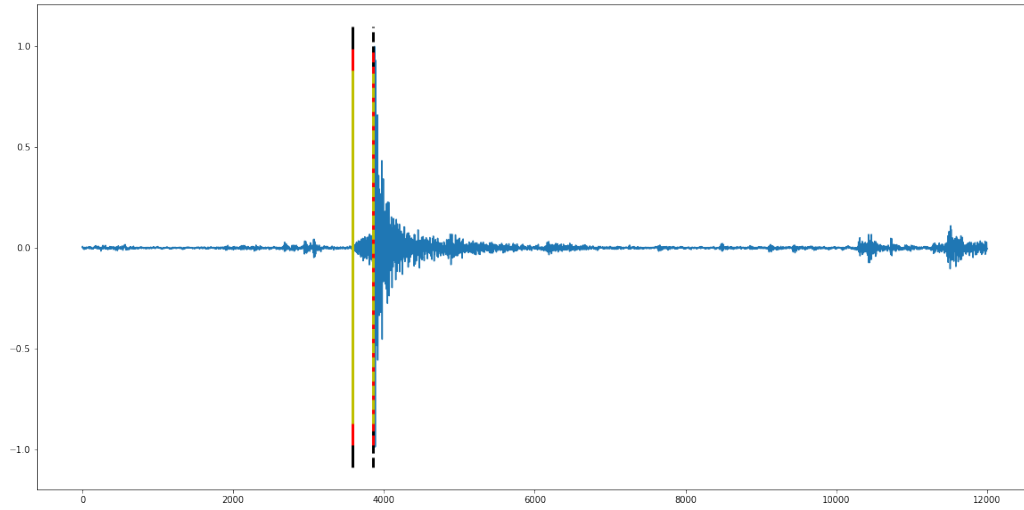
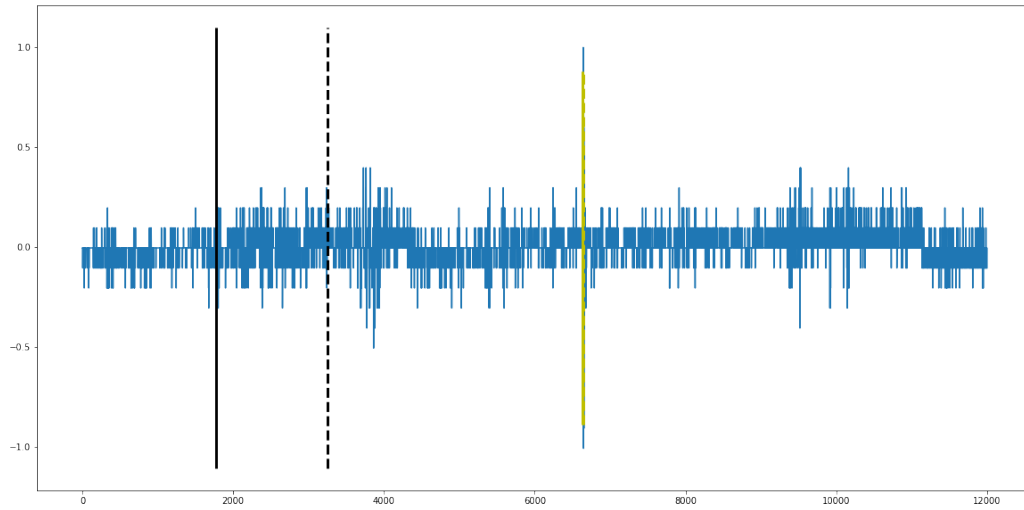(b) Almost no noise, both models work well

(c) AR Pick cheated by noise.

(d) AR Pick cheated by noise.

(e) Strong noise, EQTransformer cannot find waves and AR Pick gives a wrong prediction.

(f) Distribution of peak values in three channels (of 100,000 samples). Note that the value on x-axis is increasing exponentially.

Figure 5: Strong noise, EQTransformer cannot find waves and AR Pick gives wrong predictions.

6 Discussion

Some of our ideas to solve the phase picking problem will be introduced in this section. Since the attempts to fix and finetune PhaseNet takes up too much time and effort, these ideas still remain untried and untested.

6.1 Ideas on single-trace phase picking

Transformer and Attention[17] has been widely used in sequence processing and even image classification[6]. EQTransformer and some other recent models[10, 11] also introduces several Attention blocks into their net. Transformer is proved to be a strong tool, and we notice that in many researching fields, convolution-free and Transformer-only architectures have been raised and can rank top[4, 7, 9, 18]. We do not find similar works in the field of seismic phase picking and we wonder what a pure transformer net will bring to this task.

6.2 Ideas on multi-trace phase picking

An important concern is that in manual phase picking process, humans would use waveform data from multiple stations to determine whether an ambiguous case should be mapped to a seismic phase. As we learned from the EdgePhase[8], we may use Graph to illustrate the relation between stations, like, determine stations as neighbors if their distance is less than 100 km, and if neighbors consistently get a noise-like wave, the probabilities that it may be from a seismic phase are higher. Also we need to use single trace network to get the *ambiguous* information about the wave we detect.Then we can integrate the information in the nearby area to a more precise and reliable summary.

References

[1] Takashi Akazawa. A technique for automatic detection of onset time of p-and s-phases in strong motion records. In *Proc. of the 13th World Conf. on Earthquake Engineering*. Vancouver, Canada, 2004. 1, 2

[2] Rex V Allen. Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, 68(5):1521–1532, 1978. 1

[3] Moritz Beyreuther, Robert Barsch, Lion Krischer, Tobias Megies, Yannik Behr, and Joachim Wassermann. Obspy: A python toolbox for seismology. *Seismological Research Letters*, 81(3): 530–533, 2010. 2

[4] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021. 8

[5] Guoyi Chen and Junlun Li. Cubenet: Array-based seismic phase picking with deep learning. *Seismological Society of America*, 93(5):2554–2569, 2022. 1

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 8

[7] Dazhao Du, Bing Su, Yu Li, Zhongang Qi, Lingyu Si, and Ying Shan. Do we really need temporal convolutions in action segmentation?, 2022. 8

[8] Tian Feng, Saeed Mohanna, and Lingsen Meng. Edgephase: A deep learning model for multi-station seismic phase picking. *Geochemistry, Geophysics, Geosystems*, 23(11):e2022GC010453, 2022. 1, 8

[9] Ju He, Jie-Neng Chen, Shuai Liu, Adam Kortylewski, Cheng Yang, Yutong Bai, and Changhu Wang. Transfg: A transformer architecture for fine-grained recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 852–860, 2022. 8

[10] Wu-Yu Liao, En-Jui Lee, Dawei Mu, Po Chen, and Ruey-Juin Rau. Arru phase picker: Attention recurrent-residual u-net for picking seismic p-and s-phase arrivals. *Seismological Research Letters*, 92(4):2410–2428, 2021. 8

[11] Wu-Yu Liao, En-Jui Lee, Da-Yi Chen, Po Chen, Dawei Mu, and Yih-Min Wu. Red-pan: Real-time earthquake detection and phase-picking with multitask attention network. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. 8

[12] Alberto Michelini, Spina Cianetti, Sonja Gaviano, Carlo Giunchi, Dario Jozinović, and Valentino Lauciani. Instance–the italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021. 1

[13] S Mostafa Mousavi, William L Ellsworth, Weiqiang Zhu, Lindsay Y Chuang, and Gregory C Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1):1–12, 2020. 1, 5

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 3

[15] Tetsuo Takanami and Genshiro Kitagawa. A new efficient procedure for the estimation of onset times of seismic waves. *Journal of Physics of the Earth*, 36(6):267–290, 1988. 2

[16] Tetsuo Takanami and Genshiro Kitagawa. Estimation of the arrival times of seismic waves by multivariate time series model. *Annals of the Institute of Statistical mathematics*, 43(3): 407–433, 1991. 1

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/ 3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. 8

[18] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 568–578, 2021. 8

[19] Weiqiang Zhu and Gregory C Beroza. Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019. 1, 4