



OPEN ACCESS

EDITED BY
Tariq Alkhalifah,
King Abdullah University of Science and
Technology, Saudi Arabia

REVIEWED BY
Jiefu Chen,
University of Houston, United States
Motaz Alfarraj,
King Fahd University of Petroleum and
Minerals, Saudi Arabia

*CORRESPONDENCE
Chong Ming Wang,
✉ 1352036380@qq.com
Xing Jian Wang,
✉ wangxj@cdut.edu.cn

SPECIALTY SECTION
This article was submitted to Solid Earth
Geophysics,
a section of the journal
Frontiers in Earth Science

RECEIVED 27 October 2022
ACCEPTED 28 December 2022
PUBLISHED 10 January 2023

CITATION
Wang CM, Wang XJ, Chen Y, Wen XM,
Zhang YH and Li QW (2023), Deep learning
based on self-supervised pre-training:
Application on sandstone
content prediction.
Front. Earth Sci. 10:1081998.
doi: 10.3389/feart.2022.1081998

COPYRIGHT
© 2023 Wang, Wang, Chen, Wen, Zhang
and Li. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Deep learning based on self-supervised pre-training: Application on sandstone content prediction

Chong Ming Wang^{1*}, Xing Jian Wang^{2*}, Yang Chen¹, Xue Mei Wen¹, Yong Heng Zhang¹ and Qing Wu Li¹

¹College of Geophysics, Chengdu University of Technology, Chengdu, China, ²State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan, China

Deep learning has been widely used in various fields and showed promise in recent years. Therefore, deep learning is the future trend to realize seismic data's intelligent and automatic interpretation. However, traditional deep learning only uses labeled data to train the model, and thus, does not utilize a large amount of unlabeled data. Self-supervised learning, widely used in Natural Language Processing (NLP) and computer vision, is an effective method of learning information from unlabeled data. Thus, a pretext task is designed with reference to Masked Autoencoders (MAE) to realize self-supervised pre-training of unlabeled seismic data. After pre-training, we fine-tune the model to the downstream task. Experiments show that the model can effectively extract information from unlabeled data through the pretext task, and the pre-trained model has better performance in downstream tasks.

KEYWORDS

RNN-recurrent neural network, self-supervised, pre-train, seismic signal, sandstone content

Introduction

With the development of deep learning, the model's capability and capacity have exploded. Aided by the rapid gains in hardware, models today can easily overfit millions of data and begin to demand more—often publicly inaccessible—labeled data. However, labeling data could be extremely costly. So, researchers have recently focused on how to use unlabeled data. In Natural Language Processing (NLP), researchers successfully archive the use of unlabeled data through self-supervised pre-training. The solutions based on autoregressive language modeling in GPT (Generative Pre-Training) (Radford et al., 2018a; Radford et al., 2018b; Brown et al., 2020) and masked autoencoding in BERT (Bidirectional Encoder Representations from Transformer) (Devlin et al., 2019) are conceptually simple: they remove a portion of the data and learn to predict the removed parts (He et al., 2021). In computer vision, researchers achieve self-supervised training through contrast learning (Bachman et al., 2019; Hjelm et al., 2019; He et al., 2020). After being inspired by NLP, researchers utilized masked autoencoder with the idea of target reconstruction (He et al., 2021).

There are already some works based on self-supervised learning in seismic processing and interpretation. Such as seismic denoising, some researchers achieve random noise suppression based on Noise2Void (Krull et al., 2019), a kind of self-supervised method (Birnie et al., 2021; Birnie and Alkhalifah, 2022), some researchers achieve denoising through self-supervised seismic reconstruction (Meng et al., 2022), and some researchers achieve it by adding additive signal-dependent noise to the original seismic data and learn to predict the original data (Wu et al., 2022). What's more, self-supervised learning can also be used to reconstruct seismic data

with consecutively missing traces (Huang et al., 2022), reconstruct the low-frequency components of seismic data (Wang et al., 2020), predict facies and other properties (Zhangdong and Alkhalifah, 2020), and perform other seismic processing tasks, like velocity estimation, first arrival picking, and NMO (normal moveout) (Harsuko and Alkhalifah, 2022). These works use this method for the same reason: to solve the problem of lacking labeled data, and good results have been achieved.

However, in seismic lithology prediction, self-supervised pre-training is not widely used. Seismic lithology prediction is mainly based on supervised learning at present. Current researches are basically about the improvement of model architecture, model input, and model output. For example, using time-frequency maps of seismic data to train a convolutional neural network to predict sandstone (Zhang et al., 2018; Zhang et al., 2020), using dozens of seismic attributes to train a convolutional recurrent network to predict lithology (Li et al., 2021), and using seismic data, 90° phase-shift data, reservoir discontinuous boundary attribute to train a Multilayer Perceptron (MLP) to indirectly predict lithology by predicting cross-well natural gamma (Xin et al., 2020).

Our research is based on masked autoencoding, which is a form of more general denoising autoencoding (He et al., 2021). According to the idea of target reconstruction, the core of masked autoencoding, we design a masked autoencoder for seismic data with reference to BERT and MAE. Our model randomly masks a part of the seismic signal and reconstructs the masked part. It is based on bidirectional Long Short-Term Memory (Bi-LSTM) because the seismic signal is sequential and Bi-LSTM is good at extracting information from the sequential signal. In addition, it has an asymmetric encoder-decoder design, our decoder is lightweight compared to our encoder. We also try different masking ratios to find the best setting.

We experimented with synthetic seismic signals. After pre-training, we fine-tuned our model to the downstream task—sandstone context predicting. The results show that the pre-trained model performs better. In the downstream task, the accuracy of the pre-trained model is at best 5% higher than that of the random initialization model.

There is a lot of unlabeled seismic data in oil and gas exploration, we hope that the application of unlabeled seismic data can make deep learning better applied in seismic interpretation.

Approach

Datasets

Training and validating a Recurrent Neural Network (RNN) model often require a lot of signals and corresponding labels. Manually labeling or interpreting seismic data could be extremely time-consuming and highly subjective. In addition, inaccurate manual interpretation, including mislabeled and unlabeled faults, may mislead the learning process (Pham et al., 2019; Wu et al., 2019; Wu et al., 2020; Alkhalifah et al., 2021; Zhang et al., 2022). To avoid these problems, we create synthetic seismic signals and corresponding labels based on the convolution model for training and validating our RNN model.

Synthetic seismic signals

According to the convolution model, we know that seismic data are formed of a seismic wavelet and reflectivity:

$$S(t) = \sum_{i=-\infty}^{\infty} w(i)r(t-i) = w(t)*r(t) \quad (1)$$

where $w(t)$ is the wavelet, $r(t)$ is the reflectivity and $S(t)$ is the synthetic seismic signal, and t means time.

A ricker wavelet is a good representation of seismic wavelets, so we chose it to generate our signals:

$$R(t) = [1 - 2(\pi ft)^2]e^{-(\pi ft)^2} \quad (2)$$

where $R(t)$ is the ricker wavelet and f is the frequency of the ricker wavelet.

The reflectivity is decided by the lithology of the stratum. We can first calculate the impedance sequence Z by density ρ and velocity v of the stratum. Thus, we can calculate reflectivity from Z :

$$r_i = \frac{Z_{i+1} - Z_i}{Z_{i+1} + Z_i} \quad (3)$$

$$Z_i = \rho_i v_i \quad (4)$$

where Z_i is the impedance, r_i is the reflectivity, ρ_i and v_i are the density and velocity of the i -th layer of the stratum.

So, our workflow of creating synthetic seismic signals, see Figure 1, is as follows: we first establish a stratum of randomly distributed sandstone and mudstone, and then we calculate the impedance sequence Z using ρ and v of the sandstone and mudstone. At last, we obtain the synthetic seismic signal using Eq. 1, and the label is the corresponding stratum.

The sampling interval of the Ricker wavelet and impedance sequence is 1 ms, so the sampling interval of the synthetic seismic signal is 1 ms too.

Self-supervised pre-training of seismic signals

Our self-supervised pre-training is based on a masked autoencoder, which performs well in NLP and computer vision. A masked autoencoder masks random parts from the input signal and reconstructs the missing part. It has an encoder that maps the input signal to a latent representation and a decoder that reconstructs the original signal from the latent representation. The workflow is shown in Figure 2.

Target reconstruction

As mentioned, a masked autoencoder masks random parts from the input signal and reconstructs them. It is important to choose how to mask random parts. In this paper, we design a method for the seismic signal with reference to Masked Language Model (MLM) in BERT (Devlin et al., 2019). It is as follows.

We randomly choose a part of the signal, and we replace the chosen part with (1) zeros 80% of the time, (2) a random part of this signal 10% of the time, and (3) keep unchanged 10% of the time.

In target reconstruction task design we follow the MLM in BERT (Devlin et al., 2019) closely. According to the ablation over different masking strategies in BERT, the best strategy of masking rates of zero masking, random masking, and no masking are 80%, 10%, and 10%, respectively. The purpose of no masking is to reduce the mismatch between pre-training and fine-tuning as the signal will not be masked during the fine-tuning stage. The mixed strategy of random masking and zero masking will help the model learn better.

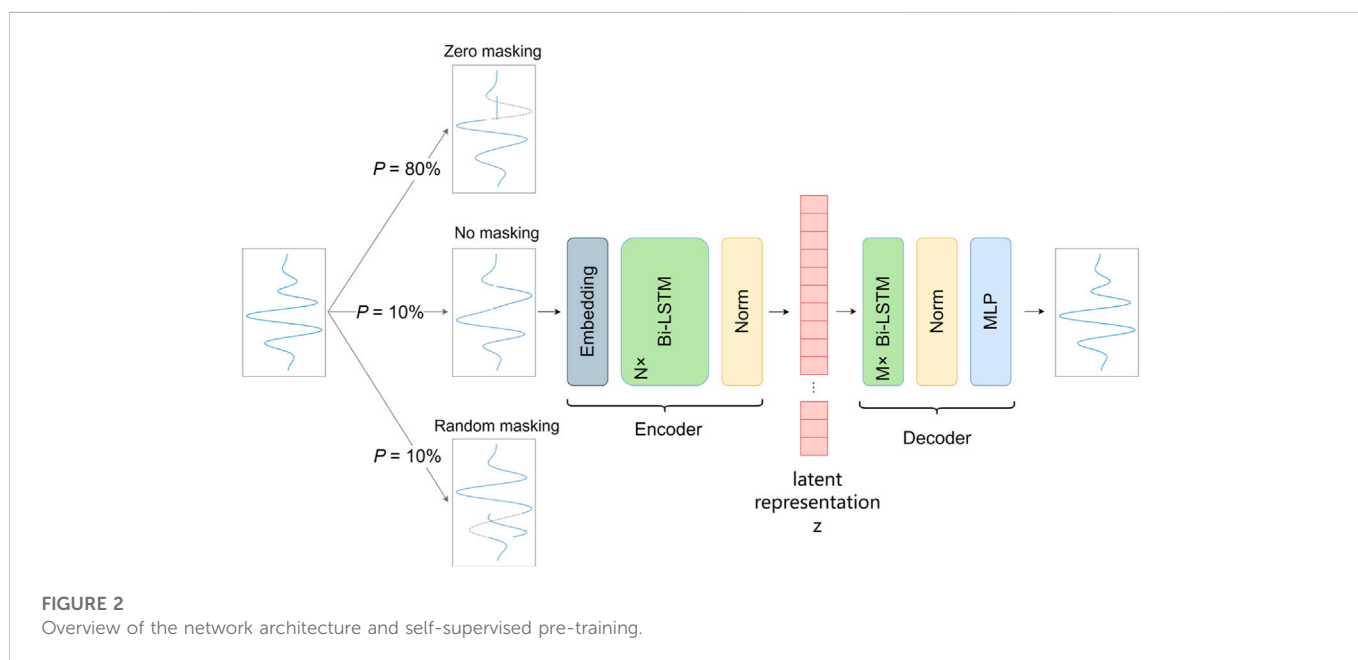
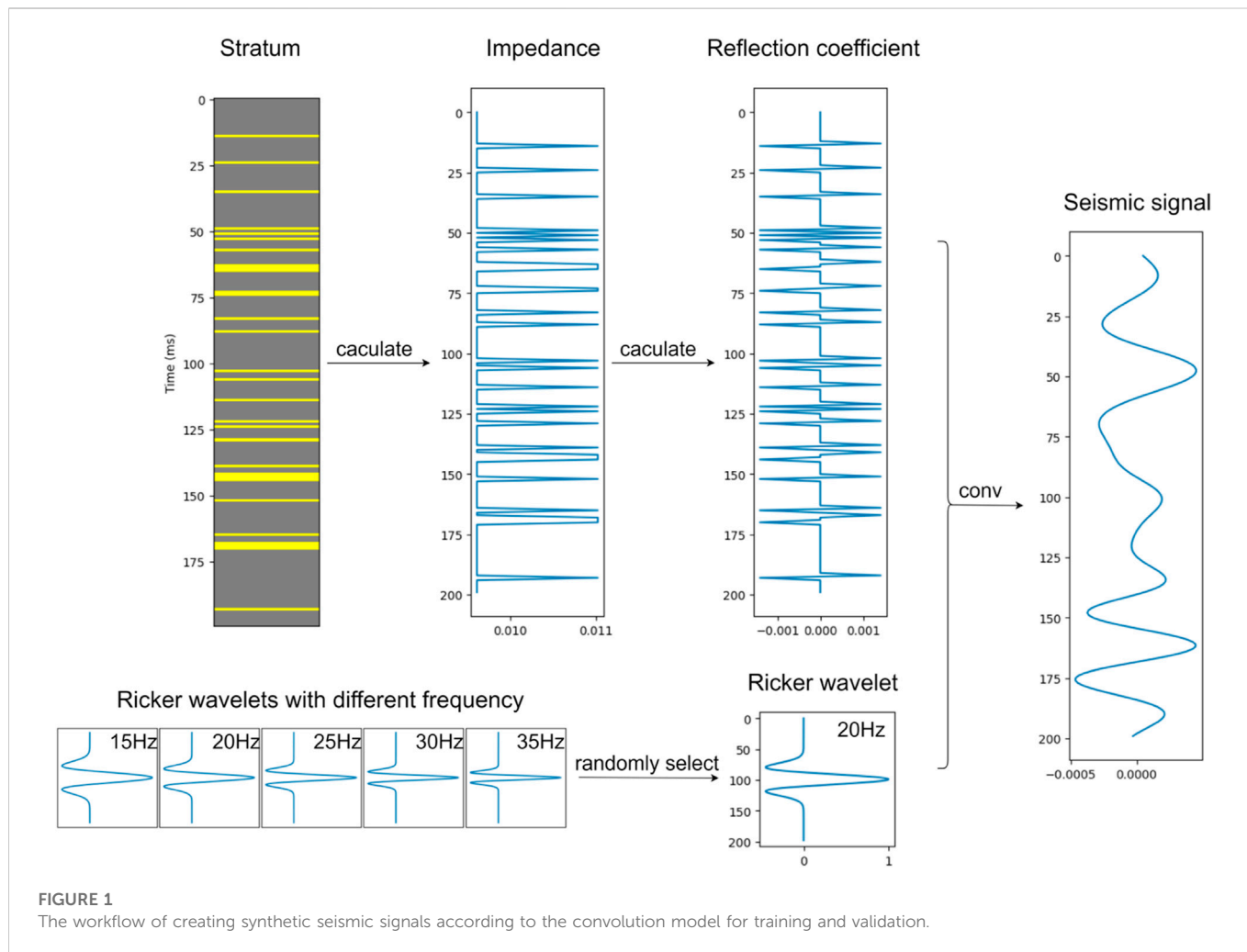


TABLE 1 Size of training and validating datasets.

	Pre-train	Fine-tune
Train	100,000	20,000
Valid	25,000	5,000

Masking ratio

Another point of the masked autoencoder is the masking ratio. It is different between different signals, when the signals are highly semantic and information-dense, the masking ratio need could be small (e.g., 10%), but on the contrary, when the signals are low in semantic, it needs to be high (e.g., 75%). For low semantic signals, a missing part can be recovered from neighboring parts without high-level understanding of signals. A high masking ratio can create a challenging self-supervisory task that requires holistic understanding beyond low-level signal statistics. We tried different masking ratios to find out the best option: about 20%.

Masked autoencoder

Masked autoencoders have an encoder and a decoder.

Encoder: An encoder maps the input signal to a latent representation. We know that a seismic signal is a sequence signal, and the RNN model is a kind of sequential model, that is good at extracting information from a sequence signal. So, our encoder is based on Bi-LSTM, a kind of bidirectional RNN. We use a bidirectional RNN because the wavelet has a width, which makes one point of the seismic signal is affected by the above and below points.

Decoder: A decoder reconstructs the original signal from the latent representation. It plays a different role in reconstructing different signals. When the decoder's output is of a lower semantic level such as pixels, the decoder needs to be designed carefully. This is in contrast to language, where the decoder predicts missing words that contain rich semantic information. So, our decoder is based on Bi-LSTM and MLP. In addition, the decoder is only used during pre-training to perform the signal reconstruction task. Therefore, it is more lightweight than the encoder to make sure effectiveness of the encoder and reduce computation.

Experiments

Our experiment is based on synthetic seismic data. Specifically, we pre-train our model on synthetic seismic data and we fine-tune our model on synthetic seismic data and corresponding labels. The specific method of creating synthetic seismic data is described above. In our experiments, the size of the dataset is shown in Table 1.

TABLE 2 Details of the pre-training model.

Encoder Bi-LSTM layers N	Decoder Bi-LSTM layers M	Hidden size D	MLP size
4	1	256	512

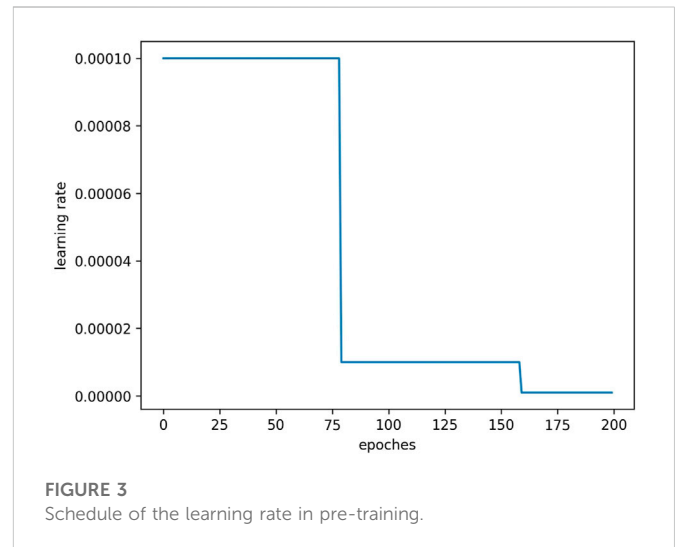


FIGURE 3
Schedule of the learning rate in pre-training.

Self-supervised pre-training

The overview of pre-training is shown in Figure 2. The N means the number of layers of the encoder's Bi-LSTM, and the M means the number of layers of the decoder's Bi-LSTM. The details of the pre-training model are shown in Table 2.

We pre-train our model using an Adam optimizer (Kingma and Ba, 2017) with $\beta_1 = .9$, $\beta_2 = .999$, and a batch size of 1,600. We use a step learning rate decay, see Figure 3. We pre-trained for 200 epochs at 4 masking ratios: .1, .2, .3, and .4. To get a qualitative sense of our reconstruction task, see Figure 4. Our loss function computes the Mean Squared Error (MSE) between the reconstructed and original seismic signals. We compute the loss only on masked parts, similar to BERT (Devlin et al., 2019) and MEA (He et al., 2021), so the model output on unmasked parts is qualitatively worse. So, we overlay the output with the unmasked parts to improve visual quality.

Fine-tuning

We can fine-tune our pre-trained model to downstream tasks, and, a new decoder for the downstream task is necessary. In this paper, the downstream task is a classification task.

Downstream task: We choose a simple task as our downstream task, so we can pay more attention to the effect of self-supervised pre-training. In the downstream task, our model predicts the sandstone content from the input seismic signal. This is a simple task but also meaningful, we can get an overall understanding of a stratum from the prediction results which can be used as a reference for further work.

The output of our downstream task is the percentage of the Sandstone content, which is suitable for regression. However, we still choose classification because a regression task is harder than a classification task. As mentioned above, we need a simple task to

TABLE 3 Details of the fine-tuning model.

Encoder Bi-LSTM layers N	Hidden size D	MLP size
4	256	1,024

validate our method. In addition, we can calculate the accuracy which can intuitively show the performance of the model.

Our model predicts the sandstone content from the input seismic signal and its decoder is a simple MLP, see Figure 5. The details of the fine-tuning model are shown in Table 3.

We fine-tune our model using again an Adam optimizer (Kingma and Ba, 2017) with $\beta_1 = .9, \beta_2 = .999$, a batch size of 2,048 and apply a weight decay of .001. We use a step learning rate decay, see Figure 6.

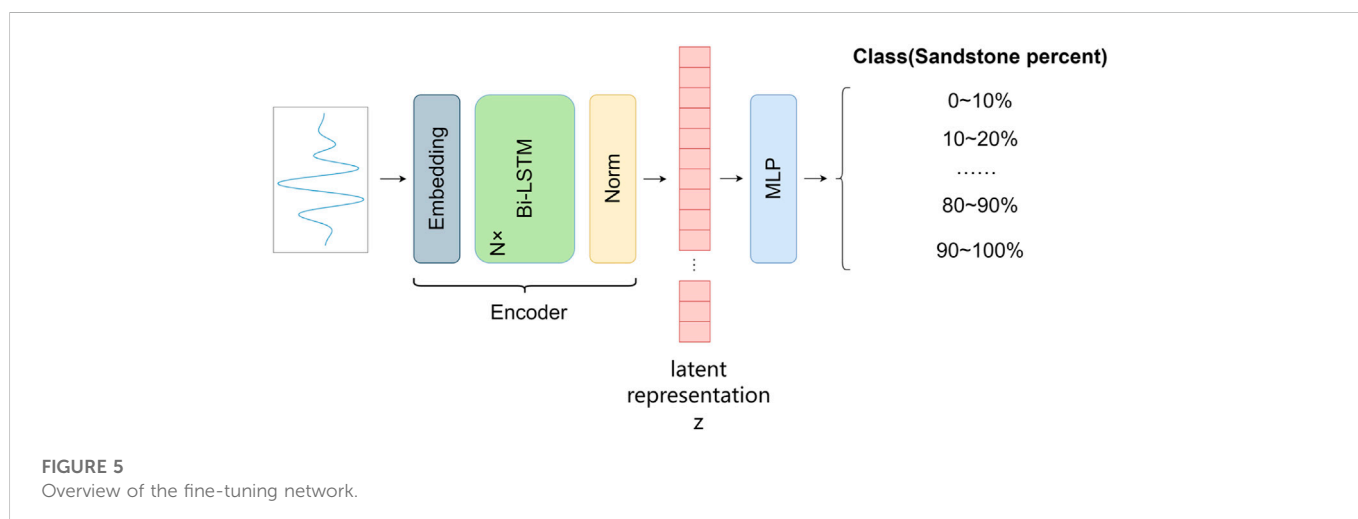
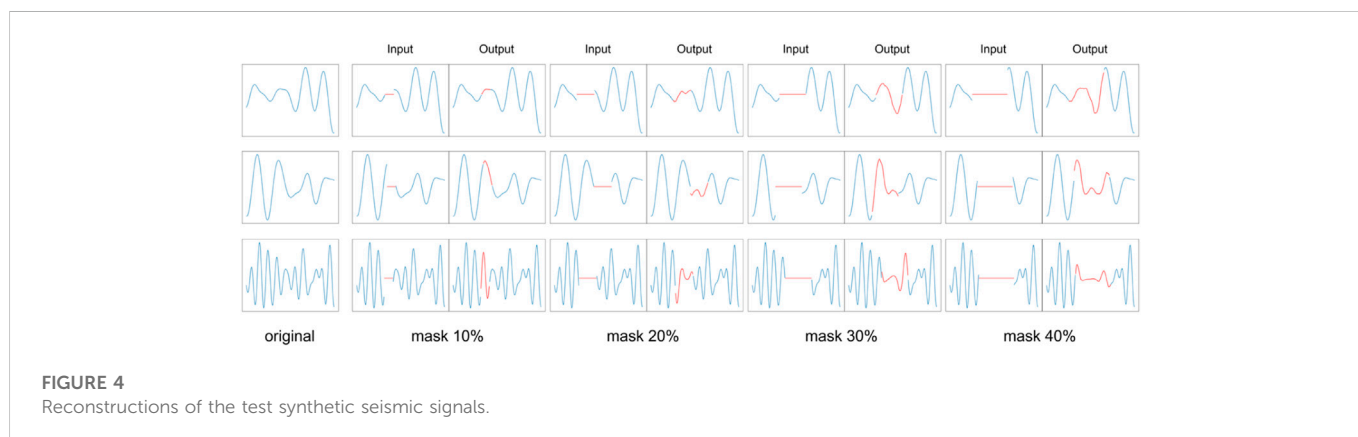
In fine-tuning we try several pre-training models with different masking ratios to find out the best choice of masking ratio. We train our model 300 epochs and record the losses during training, the losses of the last 50 epochs are shown in Figures 7, 8. We find that there is an abrupt behavior for the training and validation losses between epochs 270 and 280. We think this is caused by the an excessively high learning rate for that stage of the training. So, when the learning rate decreases, the problem disappears.

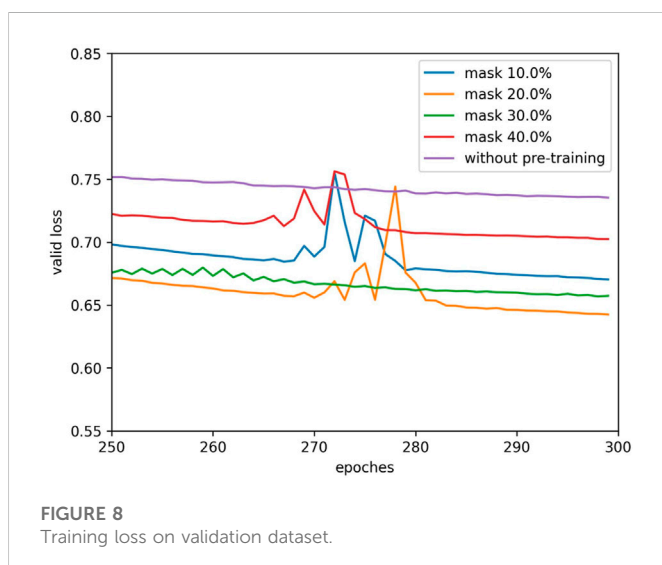
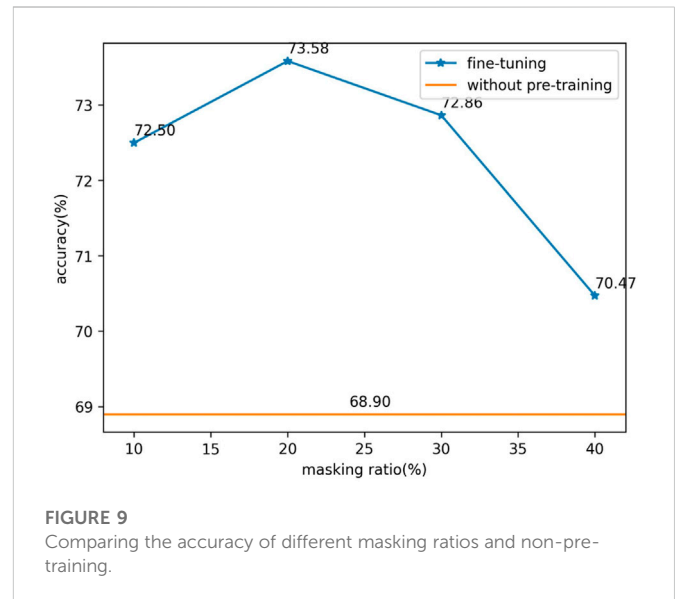
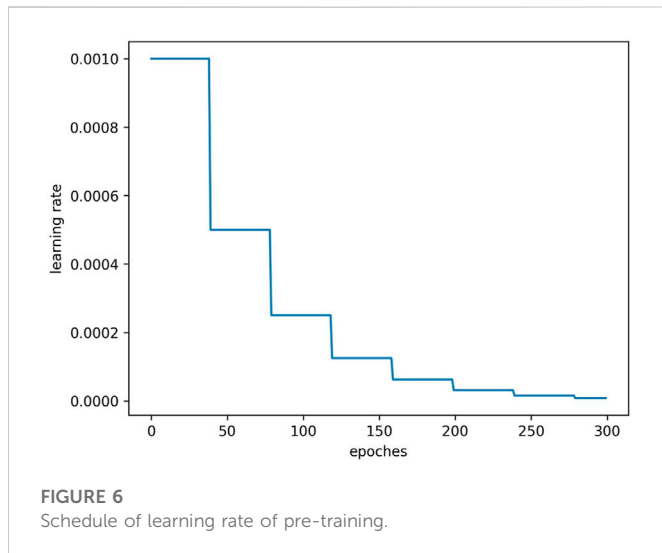
After training, we evaluate the accuracy of the models on the test synthetic dataset, see Figure 9.

Discussion

In this paper, we do successfully implement a self-supervised pre-training of seismic signals through a reconstruction task. We have tried different masking ratios and found that, in our experiment, the best choice of masking ratio is about 20%. Seismic signals are not highly semantic, a higher masking ratio should be suitable theoretically, but the truth is the exact opposite. We think it may be caused by masked input signals. We input our masked signals to the model, including the masked and unmasked parts. So, the RNN extracts information from potential noise (unmasked) and masked parts. With the rising masking ratio, the model is seriously disturbed by noise. We think a model which can input the unmasked part may solve this problem such as Transformer.

In the downstream tasks, the pre-trained model is better than the random initialization model. This proves the effectiveness of our self-supervised pre-training method for seismic signals. So, more tasks can be optimized by self-supervised pre-training in seismic deep learning.





Conclusion

With the development of deep learning, unlabeled data will be used more. On the one hand, the appetite for large-scale models is hard to address with the labeled data, because labeled data are very limited. On the other hand, extracting information from unlabeled data is an important target in artificial intelligence.

For now, self-supervised pre-training performs well in taking advantage of unlabeled data in NLP and computer vision. In our work, we designed a masked autoencoder for seismic self-supervised pre-training. When we fine-tune our model for a downstream task, the pre-trained model always performs well, which proves the effectiveness of our method. We also explore how the masking ratio influences accuracy in the downstream task.

Our model is based on RNN, which is quite costly. We think a low cost model is necessary so that we can train on bigger datasets. We think a bigger model and dataset may achieve higher accuracy. We hope that by using the self-supervised pre-training, we can better apply deep learning to seismic signals.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

CW done experiment and writing. XJW guides this work. YZ, XMW, QL and YC help CW with work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alkhalifah, T., Wang, H., and Ovcharenko, O. (2021). European Association of Geoscientists & Engineers, 1–5. MLReal: Bridging the gap between training on synthetic data and real data applications in machine learning, Proceedings of the 82nd EAGE Annual Conference & Exhibition, Amsterdam, Netherlands, October 2021
- Bachman, P., Hjelm, R. D., and Buchwalter, W. (2019). Learning representations by maximizing mutual information across views. *Adv. Neural Inf. Process. Syst.* 32. doi:10.48550/arXiv.1906.00910
- Birnie, C., and Alkhalifah, T. (2022). Transfer learning for self-supervised, blind-spot seismic denoising. <https://arxiv.org/abs/2209.12210>.
- Birnie, C., Ravasi, M., Liu, S., and Alkhalifah, T. (2021). The potential of self-supervised networks for random noise suppression in seismic data. *Artif. Intell. Geosciences* 2, 47–59. doi:10.1016/j.aiig.2021.11.001
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901. doi:10.48550/arXiv.2005.14165
- Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). Bert: Pre-Training of deep bidirectional transformers for language understanding. <https://arxiv.org/abs/1810.04805>.
- Harsuko, R., and Alkhalifah, T. (2022). StorSeismic: A new paradigm in deep learning for seismic processing. *IEEE Trans. Geoscience Remote Sens.* 60, 1–15. doi:10.1109/tgrs.2022.3216660
- He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2021). “Masked autoencoders are scalable vision learners,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, September 2021.
- He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., et al. (2020). Momentum contrast for unsupervised visual representation learning, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 2020, 9729–9738.
- Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., et al. (2019). Learning deep representations by mutual information estimation and maximization. <https://arxiv.org/abs/1808.06670>.
- Huang, H., Wang, T., Cheng, J., Xiong, Y., Wang, C., and Geng, J. (2022). Self-supervised deep learning to reconstruct seismic data with consecutively missing traces. *IEEE Trans. Geoscience Remote Sens.* 60, 1–14. doi:10.1109/tgrs.2022.3148994
- Kingma, D. P., and Ba, J. (2017). Adam: A method for stochastic optimization. <https://arxiv.org/abs/1412.6980>.
- Krull, A., Buchholz, T. O., and Jug, F. (2019). “Noise2Void - learning denoising from single noisy images,” in Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 2019, 2124–2132.
- Li, K., Xi, Y., Su, Z., Zhu, J., and Wang, B. (2021). Research on reservoir lithology prediction method based on convolutional recurrent neural network. *Comput. Electr. Eng.* 95, 107404. doi:10.1016/j.compeleceng.2021.107404
- Meng, F., Fan, Q., and Li, Y. (2022). Self-supervised learning for seismic data reconstruction and denoising. *IEEE Geoscience Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2021.3068132
- Pham, N., Fomel, S., and Dunlap, D. (2019). Automatic channel detection using deep learning. *Interpretation* 7 (3), SE43–SE50. doi:10.1190/int-2018-0202.1
- Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). *Improving Language understanding by generative pre-training*. San Francisco, California: OpenAI blog.
- Radford, A., Wu, J., Child, R., David, L., and Dario, A. (2018). *Language models are unsupervised multitask learners*. San Francisco, California: OpenAI blog.
- Wang, M., Xu, S., and Zhou, H. (2020). “Self-supervised learning for low frequency extension of seismic data,” in Proceedings of the SEG Technical Program Expanded Abstracts 2020, Virtual, May 2020, 1501–1505.
- Wu, H., Zhang, B., and Liu, N. (2022). Self-adaptive denoising net: Self-supervised learning for seismic migration artifacts and random noise attenuation. *J. Petroleum Sci. Eng.* 214, 110431. doi:10.1016/j.petrol.2022.110431
- Wu, X., Liang, L., Shi, Y., and Fomel, S. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *GEOPHYSICS* 84 (3), IM35–IM45. doi:10.1190/geo2018-0646.1
- Wu, X., Yan, S., Qi, J., and Zeng, H. (2020). Deep learning for characterizing paleokarst collapse features in 3-D seismic images. *J. Geophys. Res. Solid Earth* 125 (9), e2020JB019685. doi:10.1029/2020jb019685
- Xin, D. U., Tingen, F. A. N., and Jianhua, D. (2020). Characterization of thin sand reservoirs based on a multi-layer perceptron deep neural network. *Oil Geophys. Prospect.* 55 (6), 1178–1187.
- Zhangdong, Z., and Alkhalifah, T. (2020). High-resolution reservoir characterization using deep learning-aided elastic full-waveform inversion: The North Sea field data example. *GEOPHYSICS* 85 (4), WA137–WA146. doi:10.1190/geo2019-0340.1
- Zhang, G., Lin, C., Ren, L., Li, S., Cui, S., Wang, K., et al. (2022). Seismic characterization of deeply buried paleocaves based on Bayesian deep learning. *J. Nat. Gas Sci. Eng.* 97, 104340. doi:10.1016/j.jngse.2021.104340
- Zhang, G., Wang, Z., and Chen, Y. (2018). Deep learning for seismic lithology prediction. *Geophys. J. Int.* 215. doi:10.1093/gji/ggy344
- Zhang, G. Y., Wang, Z. Z., and Lin, C. Y. (2020). Seismic reservoir prediction method based on wavelet transform and convolutional neural network and its application. *J. China Univ. Petroleum Ed. Nat. Sci. Chin.* 44 (04), 83–93.