Vision Transformer-Based H-κ Method (HkViT) for Predicting Crustal Thickness and V_P/V_S Ratio from Receiver Functions

Youran QU **Miaosong He**
Department of Electronics Engineering and Computer Science
Peking University
{2000016625, 2000012910}@stu.pku.edu.cn

Abstract

Crustal thickness (H) and crustal V_P/V_S ratio ($\boldsymbol{\kappa}$) are fundamental parameters for regional geology and tectonics. The teleseismic receiver function (RF) is the response of the Earth structure below a seismometer to an incident seismic wave. It is commonly used to determine major interfaces of the Earth, including the above two parameters. [12] Up to now, a deep learning-based H-κ Method (HkNet) has been proposed with higher accuracy and more stable results comparing to H-κ-c method. [8] However, there are still quite large room for HkNet to improve its accuracy and robustness. Here we propose a new method which uses Vision Transformer (ViT) to estimate H and κ. [3] Our model can be divided into two parts. The first part is set to denoise the receiver functions, while the second part uses ViT to predict H and κ with denoised RFs being its input. Synthetic data tests and real data both show that our new method obtains great accuracy and robustness.

1 Introduction

Crustal thickness (H) and crustal P-wave and S-wave velocity ratio ($\boldsymbol{\kappa}$) are fundamental parameters which can provide an essential basis for regional tectonics and dynamics. However, due to the complexity of the crustal structure, we can only obtain these parameters from calculating instead of measuring which brings unnecessary while heavy work to researchers. The teleseismic RF technique is the response of the Earth structure below a seismometer to an incident seismic wave. It is treated as an efficient seismic tool for imaging discontinuities at crustal and upper mantle depths beneath broadband seismic stations. [7, 9, 1, 11]

The idea of estimating H and κ from RFs has been put forward for decades. The H-κ method developed by Zhu and Kanamori (2000) has been widely used to estimate the thickness and average V_P/V_S of the crust through a grid search of H and κ values. [14] Several years ago, Li (2019) proposed a new method that corrected for the effect of crustal anisotropy and dipping interface on RFs, which results in more reliable estimation of crustal thickness and V_P/V_S ratio. [8] Last year, a deep learning-based method was developed by Wang (2022) which can automatically extract information from RFs. Compared with traditional H-κ or the recently developed H-κ-c methods, their method has higher accuracy and more stable results, as demonstrated through extensive and systematic synthetic tests. [12]

Our goal is to build a model to predict H and $\boldsymbol{\kappa}$ from a receiver function(RF) and improve its effectiveness and robustness to the utmost extent.

The contributions of our work are listed as follows:

1. We propose the idea of combining Vision Transformer with a geological estimating task, and successfully build the model as expected. This is the first attempt on using ViT to estimate crustal thickness and crustal P-wave and S-wave velocity ratio.

2. Instead of experimenting on a single real data piece, we use a large amount of real data to conduct experiments and obtain their average errors with the results from other traditional methods.

3. We analyze the denoise net's impact on estimating synthetic data and real data and their differences with the demonstration of experiments' results.

In this project, Youran proposed the theoretical framework, did the main coding part and polished the report. Miaosong wrote the main body of the report and slides and helped conduct experiments.

Code and Data are available on https://github.com/Madscientist833/HKViT

2 Related Work

Estimating crustal thickness and V_P/V_S ratio from RFs In Geology and Geophysics, estimating crustal thickness and V_P/V_S ratio from RFs is considered as the most general and effective method. The H-κ method put forward by Zhu and Kanamori (2000) has been diffusely used to predict the thickness and average V_P/V_S of the crust through a grid search of their values. [14] Li (2019) developed a new method that corrected for the effect of crustal anisotropy and dipping interface on RFs, which results in more accurate and reliable estimation. [8] Wang (2022) proposed a method for automatically estimating H and κ from RFs based on deep learning and neural network.

3 Method

3.1 Workflow

The workflow of our model is shown in Figure 1. We choose simulated RF or real data as the input of our model. Denoise net is an optional choice. Denoised RF or Simulated RF will be converted to 3-channel image, and resized from 500*73 to 224*224. The image is then send to HkViT or different kinds of CNNs to estimate the values of H and κ.

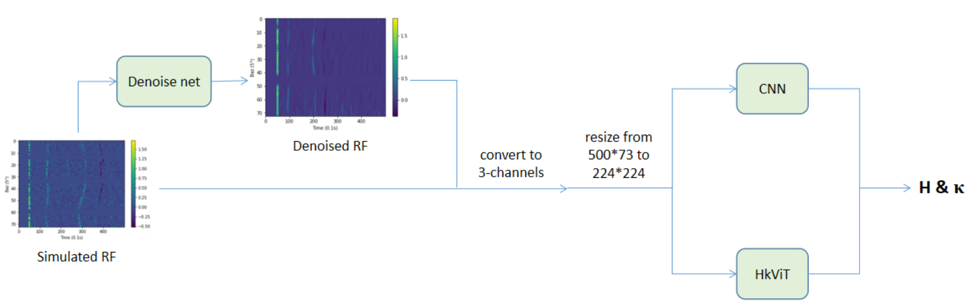
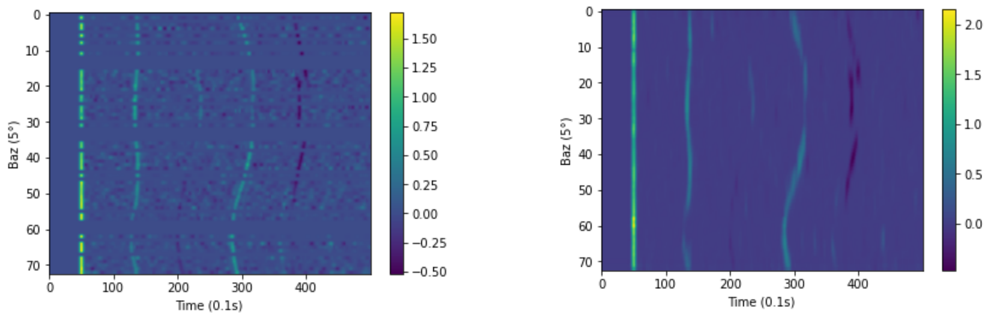


Figure 1: **Illustration of workflow**

3.2 Denoise net

After a full-scale view of all data, we find out that real data contains much more noise comparing to original data because of monitors or other reasons. To ensure our models' robustness under real data, we can use synthetic data to stimulate noise or train a denoise net from synthetic and original data to denoise real data.

Many traditional denoise methods have been developed during the history of computer science (eg. Gaussian filter); In recent years, deep learning-based methods have been put forward and proved to be more efficient methods comparing to tradition methods. [13] Thus we propose a Denoise net with simple CNN architecture to handle the noise in real data. The network is composed by 4 convolutional layers which is rather simple, but it seems to be efficient enough in experiments.(PSNR[2]:28.90)



(a) **Original simulated RF** (b) **Denoised RF**

Figure 2: **The comparison of original simulated RF and denoised RF**

3.3 Data preprocessing

After the denoising part has been finished, the simulated RF or denoised RF will be converted to a 3-channel image and resized from 500*73 to 224*224.Then we normalized these image with [0.485, 0.456, 0.406] as mean; [0.229, 0.224, 0.225] as std. This series of operations are set to make best use of the weights of pre-training which play an important role in accelerating converging process. The training process might not even be proceeded if data preprocessing part is missing which can be demonstrated by experiment results.

3.4 Predicting network structure

For the estimating task, we take receiver function as input (3*224*224 after preprocessing) which could be considered as a 3-channel image. and the output of the network are the values of H and κ. Therefore, it is natural to consider CNN as the network structure for the regression task. We have tested some CNN architectures including VGG[10], ResNet[5], and there are plenty of choices to be experimented in future work.

Besides CNN, transformers have grown more and more popular in vision tasks [3], and reach SOTA performance in almost all tasks. Specifically, in fine-grained classification task, almost all methods are based on transformers since 2021, which have been demonstrated to be more efficient than CNN. [4] In our project, differences between images are quite small which makes our model hard to precisely predict the values of H and κ. Apart from that, as shown in Figure 3, the curves highlighted by red rectangles in the image are crucial areas thich determine the values of H and κ. According to these traits of our task mentioned above, it is obvious that our task shares a lot of similarities with fine-grained classification task, so we try ViT as an option for predicting network as well.

To improve the estimating results of H and κ, we have trained our predicting network for H and κ separately instead of jointly. With this strategy, the model's loss can focus on a single target, which ensure the weights of the model to reach the best fit for each target.

To enhance the model‘s performance and the efficiency of training process, we use the weights pretrained on ImageNet [6] for almost each architecture, and it appears that models with pretrained weights outperform models without pretrained weight.

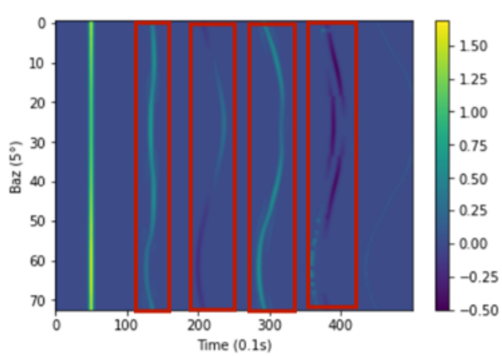


Figure 3: **Crucial features in receiver function for estimating H and κ**

4 Experiment

In the training process, we designated the same hyperparameter setting for each model in order to compare performances between models fairly. For denoise net, we set learning rate=1e-3; batchsize=4; optimizer=SGD; epoch=20; scheduler=Linearscheduler with warmup. For predicting task, we set learning rate for predicting H as 1e-5, predicting κ as 1e-3; batchsize=4; optimizer=SGD; epoch=10; scheduler=Linearscheduler with warmup.

4.1 Synthetic Tests

We first used the simulated RFs as the input of our network to estimate H and κ, We tried both predicting H and κ directly and predicting them after applying denoise net. The results are listed in Table 1.

Table 1: **Errors of different network structures from synthetic tests**

Model Name	Train data		Test data	
	dH/H(%)	dκ/κ(%)	dH/H(%)	dκ/κ(%)
Resnet18	11.65	4.5	12.55	4.5
Resnet18_denoised	**2.6**	1.8	**3.2**	2.5
ViT	10.91	4.2	14.53	4.2
ViT_denoised	5.8	2.1	5.7	3.4
vgg16	-	4.9	-	5.0
vgg16_denoised	-	1.6	-	2.0
vgg19	-	4.6	-	4.6
vgg19_denoised	-	**1.5**	-	**1.8**

- means loss is so huge that the model cannot be trained.

As shown in Table 1, Resnet18 with denoise net and Vgg19 with denoise net reached the best performance on dH/H and dκ/κ in both train data and test data among all models. The reasons why these CNN models outperformed ViT were considered as follows: Our training steps were too small; Deep and large models such as ViT have not been trained sufficiently. Additionally, we found that models with denoise net outperformed raw net, which might due to synthetic data are more smooth thus easier to fit.Overall, our model reached a very high performance on test data with dκ/κ(%)=**1.8** and dH/H(%)=**3.2**

4.2 Application to Real Data

Then we tested our model on real data. The results are listed in Table 2.

As shown in Table 2, Resnet18 without denoise performed best on dH/H, and ViT without denoise performed best on dκ/κ. We recognized that unlike synthetic data, models without denoise performed better than models with denoise in real data situation. After having a thorough research of some specific features of real data, we found out that noise of real data were much more complicated than our training data, so our denoise net were not efficient enough to handle real data. Besides, ViT performed best on dκ/κ, which showed the efficiency of self-attention mechanism in complicated cases. Lastly, errors in dκ are much less than dH, which might give reason to real data's H varies a lot while κ changed in a small range. Overall, our model perform well for predicting κ(dκ/κ(%)=**2.4**), but for H, our model didn't handle it very well.

5 Conclusion

In this work, we devise and realize a complete workflow to estimate crustal thickness (H) and P-wave and S-wave velocity ratio (κ) from receiver functions. The punchline of our work is bringing in

Table 2: **Errors of different network structures from real data tests**

Model Name	dH/H(%)	dκ/κ(%)
Resnet18	**19.59**	3.1
Resnet18_denoised	27.36	5.8
ViT	21.69	**2.4**
ViT_denoised	21.04	4.2
vgg16	-	3.4
vgg16_denoised	-	4.6
vgg19	-	3.3
vgg19_denoised	-	4.2

- means loss is so huge that the model cannot be trained.

Vision Transformer as the module to predict the values of H and κ whose conspicuous performance and robustness can be demonstrated by the experiment results in both synthetic and real data tests. In addition, we find out that denoise net seems to have negetive effects on estimating tasks of real data.

The contributions of our work are listed as follows:

1. We propose the idea of combining Vision Transformer with a geological estimating task, and successfully build the model as expected. This is the first attempt on using ViT to estimate crustal thickness and crustal P-wave and S-wave velocity ratio.
2. Instead of experimenting on a single real data piece, we use a large amount of real data to conduct experiments and obtain their average errors with the results from other traditional methods.
3. We analyze the denoise net's impact on estimating synthetic data and real data and their differences with the demonstration of experiments' results.

6 Future Work

Due to multiple complicated reasons, time for us to proceed this project seems a bit tight. Thus, we have quite a lot ideas to be realized and they are listed below.

1. We'll conduct experiments with more steps and various settings of hyperparameters for the experiments' results have not reached the preconceived level up to now.
2. We'll apply object detection on curves highlighted in Figure 3 and seperately extract features which are useful for estimating.
3. As the denoise net shows no obvious capacity in reducing the estimating errors of H and κ, it's crucial to construct a well-designed and more effective denoise net.
4. We'll manage to apply network architecture designed for fine-grained-classification task.

References

[1] Charles J Ammon. The isolation of receiver effects from teleseismic p waveforms. *Bull. Seismol. Soc. Am*, 81(6):2504–2510, 1991. 1

[2] Gisle Bjontegaard. Calculation of average psnr differences between rd-curves. *VCEG-M33*, 2001. 2

[3] 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. 1, 3

[4] 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. 3

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 3

[7] Charles A Langston. Structure under mount rainier, washington, inferred from teleseismic body waves. *Journal of Geophysical Research: Solid Earth*, 84(B9):4749–4762, 1979. 1

[8] Jiangtao Li, Xiaodong Song, Pan Wang, and Lupei Zhu. A generalized h-κ method with harmonic corrections on ps and its crustal multiples in receiver functions. *Journal of Geophysical Research: Solid Earth*, 124(4):3782–3801, 2019. 1, 2

[9] Thomas J Owens and George Zandt. The response of the continental crust-mantle boundary observed on broadband teleseismic receiver functions. *Geophysical Research Letters*, 12(10): 705–708, 1985. 1

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3

[11] Timur Tezel, Takuo Shibutani, and Bülent Kaypak. Crustal thickness of turkey determined by receiver function. *Journal of Asian Earth Sciences*, 75:36–45, 2013. 1

[12] Feiyi Wang, Xiaodong Song, and Jiangtao Li. Deep learning-based h-κ method (hknet) for estimating crustal thickness and vp/vs ratio from receiver functions. *Journal of Geophysical Research: Solid Earth*, 127(6):e2022JB023944, 2022. 1

[13] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2

[14] Lupei Zhu and Hiroo Kanamori. Moho depth variation in southern california from teleseismic receiver functions. *Journal of Geophysical Research: Solid Earth*, 105(B2):2969–2980, 2000. 1, 2