Bulletin of Earth Sciences of Thailand

Performance Analysis of SegFormer for Fault Prediction

Ekapost Meenak[1*] and Pongthep Thongsang[1]

[1] M.Sc. Petroleum Geoscience Program, Department of Geology, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand

*Author e-mail: EkapostM@pttep.com

Abstract

Accurately identification of faults in seismic images is critical in reservoir characterization, structural geological interpretation, and well placement. While traditional methods rely on horizontal discontinuities in seismic reflectivity, they are often plagued by artifacts and require manual correction as post-processing. SegFormer, based on a backbone transformer, has successfully tackled this challenge by utilizing the open-source dataset from the Thebe gas field located in Australia's Carnarvon Basin. The experiment involved comparing three primary backbone models, namely, MiT-B1, MiT-B3, and MiT-B5, along with hyperparameter tuning. The tests demonstrate that the lowest cross-entropy loss is MiT-B5 with 98.8% accuracy. In addition, it is important to highlight that SegFormer demonstrates excellent accuracy in predicting faults during 3D post-stack seismic migration and generates minimal artifacts. This suggests that the technology has the potential to supplant traditional seismic attributes in fault interpretation processes, indicating its promising role in fault analysis.

Keywords: SegFormer, Machine learning, Deep learning, Transformers, Seismic interpretation

1. Introduction

Faults are geological features that form in the upper crust of the Earth because of brittle deformation (Fossen, 2010). Accurately mapping faults is crucial for predicting the distribution and size of natural resources and mitigating geohazard risks (Richards et al., 2015; Fossen, 2010). Traditionally, fault mapping was conducted through the examination of natural outcrops or mines (Lisle, 2004). However, due to remote or poorly exposed study areas, remote-sensing techniques, and borehole data, such as well data, have become increasingly utilized in fault mapping (Csillag and Stogicza, 1987). The most significant advancement in recent times for fault mapping is the widespread availability of 3D seismic reflection surveys. These surveys provide detailed images of extensive rock volumes and the intricate 3D fault networks within them. A seismic reflection survey generates a subsurface image by detecting density contrasts between different rock layers, resulting in continuous reflections at their interfaces. Faults, although not directly imaged, can be identified by the disruptions they cause in these continuous reflections. Manual digitization of faults, represented as lines where multiple reflections are offset, forms a significant part of subsurface interpretation based on seismic data (Biondi et al., 2007; Robein, 2010). Traditionally, the interpretation of faults from seismic reflection data has been a manual process relying on the expertise and experience of interpreters (Gibson et al., 2012; Alcalde et al., 2019). While fault mapping from high-quality seismic data is relatively objective, it is a time-consuming and repetitive task, often taking weeks to months to complete the interpretation of a single dataset (Gibson et al., 2012). Despite the use of good quality seismic datasets, there are often signal disturbances in the data, especially in areas with numerous faults. These disturbances can lead to poorly resolved discontinuities in seismic reflectors, resulting in imprecise fault locations or misinterpretations (Badley et al., 1991; Iacopini et al., 2016; Alcalde et al., 2019). Consequently, the interpretation results may have a higher level of uncertainty, which is frequently overlooked

and challenging to quantify. Meanwhile, the rapid growth in seismic data acquisition has created challenges associated with big data for many oil and gas companies and research groups (Mohammadpoor and Torabi, 2018). Associated increasing demands for automatic or semi-automatic seismic fault interpretation reflect geologists' and geophysicists' requirements within the oil and gas industry. In some early work, researchers were mainly focused on creating and utilizing several derivative volumes and attributes to supplement the standard fault interpretation workflow, including dip and azimuth, coherence (Marfurt et al., 1999), curvature (Roberts, 2001), variance (Silva et al., 2005) and semblance (Marfurt et al., 1998). Recent technological advancements have led to semi-automatic computer solutions that utilize basic feature analysis techniques such as Ant Tracking (Silva et al., 2005) and Hough transform (Wang and AlRegib, 2014). These techniques have shown great potential in addressing various challenges in the field. However, because most of those methods are sensitive to noise (Marfurt et al., 1998; Silva et al., 2005; Cohen et al., 2006; Yan et al., 2019), seismic datasets often need to be preprocessed to a certain signal-to-noise level (Yuan et al., 2019), which in view of the unique geological structure and data quality of different datasets, it is usually necessary to manually adjust parameters by trial and error.

Previous methods in fault interpretation often relied on manual selection of a few features, which introduced subjectivity into the process. To overcome this limitation, machine learning techniques involving multiple features, such as support vector machine (SVM) and multilayer perceptron (MLP), have been explored to achieve more accurate interpretation results (Kortström et al., 2016; Di et al., 2017; Guitton et al., 2017; Di et al., 2019). Although these methods have had a positive impact on minimizing the need for manual intervention, there is still ample opportunity for progress in terms of accuracy, efficiency, and ease of use. Advancements must be made to effectively achieve these objectives.

In the past decade, deep convolutional neural networks (DCNNs) have emerged as a promising approach that surpasses human performance in various computer vision tasks, as demonstrated by the well-known ImageNet large-scale visual recognition challenge (Russakovsky et al., 2014). In recent years, there have been advancements in applying deep learning algorithms to fault interpretation, but these efforts are still in the early stages of development. Most of the current research utilizes synthetic data, disregarding the complex geological structures and noise present in real field data (Wu et al., 2019). As a result, manual interpretation of seismic data remains the dominant workflow in the industry and academia.

Recently, some convolutional-neural-network (CNN) methods have been introduced to detect faults by pixel-wise fault classification (fault or nonfault) with multiple seismic attributes (Huang et al., 2017; Di, 2018; Guitton, 2018; Guo et al., 2018; Zhao and Mukhopadhyay, 2018). Wu et al. (2018) use a CNN-based pixelwise classification method to not only predict the fault probability but also estimate the fault orientations at the same time. These methods need to choose a local window or cube to make fault prediction at every image pixel, which is computationally highly expensive, especially in 3D fault detection.

The research aims to accomplish two primary objectives. Firstly, this study proposes a solution to address the class imbalance issue in fault binary images by leveraging the SegFormer model. The SegFormer model is specifically selected due to its exceptional capability to capture global contextual information and comprehend spatial relationships, rendering it highly suitable for

fault prediction tasks in geology, oil exploration, and earthquake risk assessment. Through the integration of attention blocks and the combination of transformer-based attention mechanisms with CNN-based feature extraction, the SegFormer model demonstrates its proficiency in comprehensively analyzing seismic signals and accurately identifying intricate fault patterns. Additionally, the model utilizes lightweight multilayer perceptron (MLP) decoders to simplify the architecture and aggregate information from multiple layers, thereby facilitating the generation of robust representations. The amalgamation of local and global attention mechanisms empowers SegFormer to efficiently perform segmentation tasks using Transformers, as highlighted by Xie et al. (2021) and Vaswani et al. (2017).

Secondly, this research aims to enhance the fault recognition performance by employing a balanced cross-entropy loss function in the SegFormer model. This loss function effectively tackles the class imbalance issue that arises from the considerable number of zero-labeled pixels compared to the limited number of one-labeled pixels representing faults. By considering both positive and negative samples during the optimization process, the model is trained to accurately capture fault patterns. The widely used cross-entropy loss, commonly employed in deep learning for classification tasks, quantifies the dissimilarity between predicted probabilities and true labels. Smith et al. (2017) provide a comprehensive analysis of the cross-entropy loss, emphasizing its crucial role in training neural networks and exploring strategies for its effective application. Moreover, Li and Vasconcelos (2018) investigate the relationships between the cross-entropy loss and alternative loss functions, offering insights into their trade-offs and suitability in different contexts.

2. Methodology

In this study, our primary focus is on the pre-processing pipeline and the training/validation process employed for training the SegFormer model with a specific emphasis on fault prediction. The pre-processing pipeline consists of a series of carefully designed steps that aim at efficiently preparing the data and maximizing the model's performance.

1.Data preparation

Data Loading:

In deep learning models, datasets are typically divided into training, validation, and test sets. The training set is utilized to train the model's parameters, while the validation set assists in fine-tuning these parameters. For this study, we primarily utilized a publicly available dataset from the Thebe Gas Field in the Exmouth Plateau of the Carnarvon Basin in Australia. To investigate the structural characteristics and evolution of the basin, expert interpreters manually labeled faults from seismic data. The dataset is presented in a numpy array format, representing a seismic volume pixelated at the same resolution as the seismic data. The seismic data images depths of approximately 4.5 seconds Two-Way Travel Time (TWT), equivalent to a depth of around 3.7 kilometers. The dataset is presented in a numpy array format, representing a seismic volume pixelated at the same resolution as the seismic data. The seismic data images depths of approximately 4.5 seconds Two-Way Travel Time (TWT), equivalent to a depth of around 4 kilometers. The dataset covers an area of 45 kilometers by 39 kilometers, providing a total areal extent of 1200 square kilometers. The crosslines are spaced at 25 meters, and the inlines are spaced at 12.5 meters.

The dimensions of the numpy array are 1803[crossline] × 1537[sample] × 3174 [inline]. The sample dimension represents the vertical axis, typically representing different time slices or depth levels. In this case, there are 1537

samples, indicating that the seismic data is divided into 1537 layers or slices along the vertical direction. Each sample represents a specific depth or time slice, allowing for detailed analysis of the subsurface structure. To maintain consistency, the first four out of 1807 crosslines were excluded from the dataset due to a lack of corresponding annotations. Figure 1 illustrates the processed dataset.

Data Splitting:

Random splitting is a common approach to divide data into training and test sets in machine learning. However, for geological faults, adjacent slices often exhibit highly similar distributions. Therefore, random splitting is not appropriate as it could lead to "peeking" into the test set, resulting in an overly optimistic model. Instead, we divided the dataset into blocks in a ratio of approximately 5:1:4, corresponding to the training, validation, and test sets, respectively. Specifically, the decision to choose crosslines as part of the training set was driven by their ability to represent the geological structures in a perpendicular manner to the main fault's strike. The training set comprised the first 900 pairs of crosslines, consisting of seismic data traces and their corresponding fault masks. Additionally, the subsequent 200 pairs were allocated to the validation set, and the remaining 703 pairs constituted the test set.

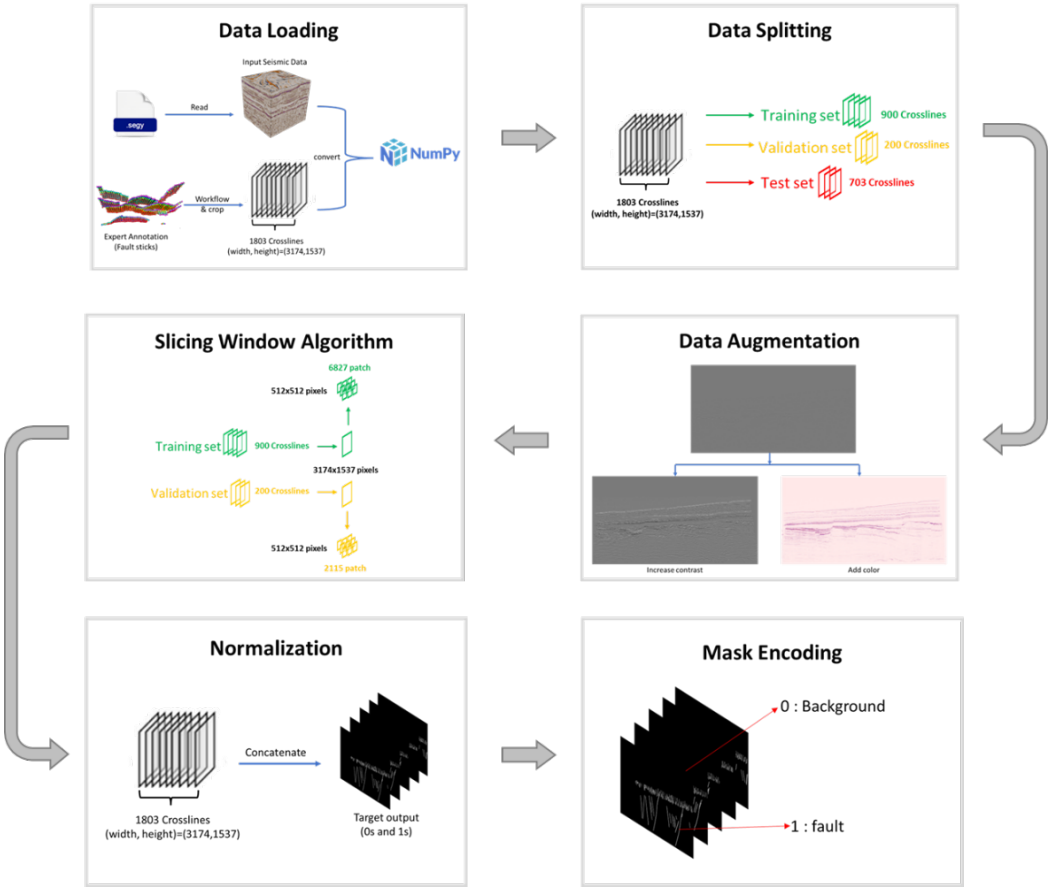

Figure 1 The procedure for data preparation

Data Augmentation:

Data augmentation techniques were employed to double the diversity of the training

dataset. and improve the model's generalization. In this study, we adjusted the contrast and added color to the seismic data using seismic data augmentation techniques.

Slicing Window Algorithm:

The geological interpretation of the Thebes dataset was conducted by expert interpreters from the Fault Analysis Group at University College Dublin (An et al., 2021). Their primary focus was on identifying faults with vertical displacements exceeding 20 m within a specific area of interest and depth range, which encompassed approximately 2 km to 4 km. The shallower and deeper parts of the seismic volume were generally disregarded. To ensure clarity and optimize Graphics Processing Unit (GPU) memory usage while generating a sufficiently large training set, we utilized a slicing window algorithm. This algorithm divided the seismic volume into patches measuring 512 × 512 pixels. Patch pairs lacking fault label or having an insufficient proportion of labeled pixels were filtered out. Reflective padding was applied to ensure a whole number of patch pairs. Consequently, the training set consisted of 6,827 patch pairs, while the validation set contained 2,115 patch pairs.

Normalization:

Normalization was performed to standardize the pixel values of the input images. In this study, we employed min-max normalization. Fault interpretations were represented by binary images for each crossline. Resolution of these images was identical to the original seismic data. Fault pixels were marked as 1, indicating the presence of a fault, while non-fault pixels were marked as 0. The processed dataset is depicted in Figure 1.

Mask Encoding:

Segmentation masks are typically represented as pixel-wise labels, where each pixel is assigned a specific class value indicating the object it belongs to. In our case, fault masks were encoded into a suitable format for training by assigning a value of 1 to faults and 0 to the background class.

2. Training and validation of fault recognition models.

The SegFormer model is a deep learning architecture specifically designed for semantic segmentation tasks (Xie et al., 2021). It combines the advantages of transformers, commonly used in natural language processing, with the effectiveness of convolutional neural networks (CNNs) in computer vision. The model architecture employs a hybrid structure that consists of a convolutional backbone followed by a transformer encoder-decoder, as shown in Figure 2. As depicted in Figure 3, SegFormer architecture consists of two main modules: a hierarchical Transformer encoder and a lightweight All-MLP decoder.

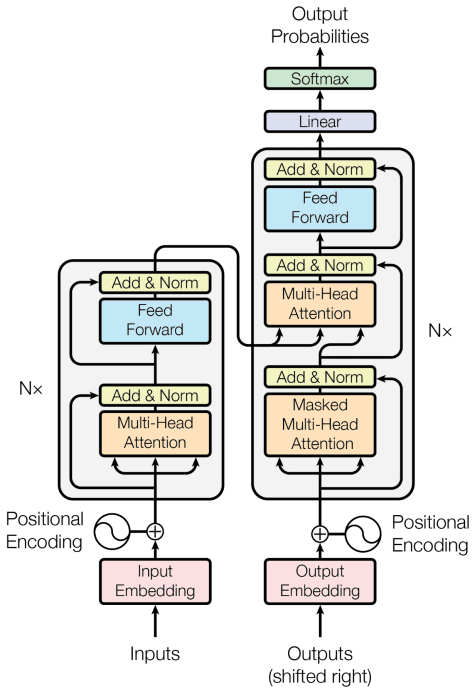

Figure 2 The Transformer – model architecture (Vaswani et al., 2017)

The transformer encoder-decoder is responsible for capturing global dependencies and generating high-resolution segmentation maps. The encoder consists of multiple transformer blocks that sequentially process the feature maps from the convolutional backbone. Each transformer block contains self-attention mechanisms that attend to different parts of the input feature maps, enabling the model to capture long-range dependencies and contextual information. The decoder uses a combination of upsampling and convolutional layers to generate the final segmentation maps.

The training process:

The SegFormer model involves several key steps to optimize its performance in fault segmentation. The dataset is prepared by dividing it into batches, each containing images and their corresponding ground truth segmentation masks. An optimization algorithm, in this study use AdamW, is utilized to update the model's parameters, with carefully tuned hyperparameters like learning rate and weight decay. The choice of loss function, such as cross-entropy or mean squared error, plays a significant role in training the model by measuring the discrepancy between predicted and ground truth segmentation maps. Model initialization techniques, including random initialization, pre-training on ImageNet, or transfer learning, are employed to enhance performance.

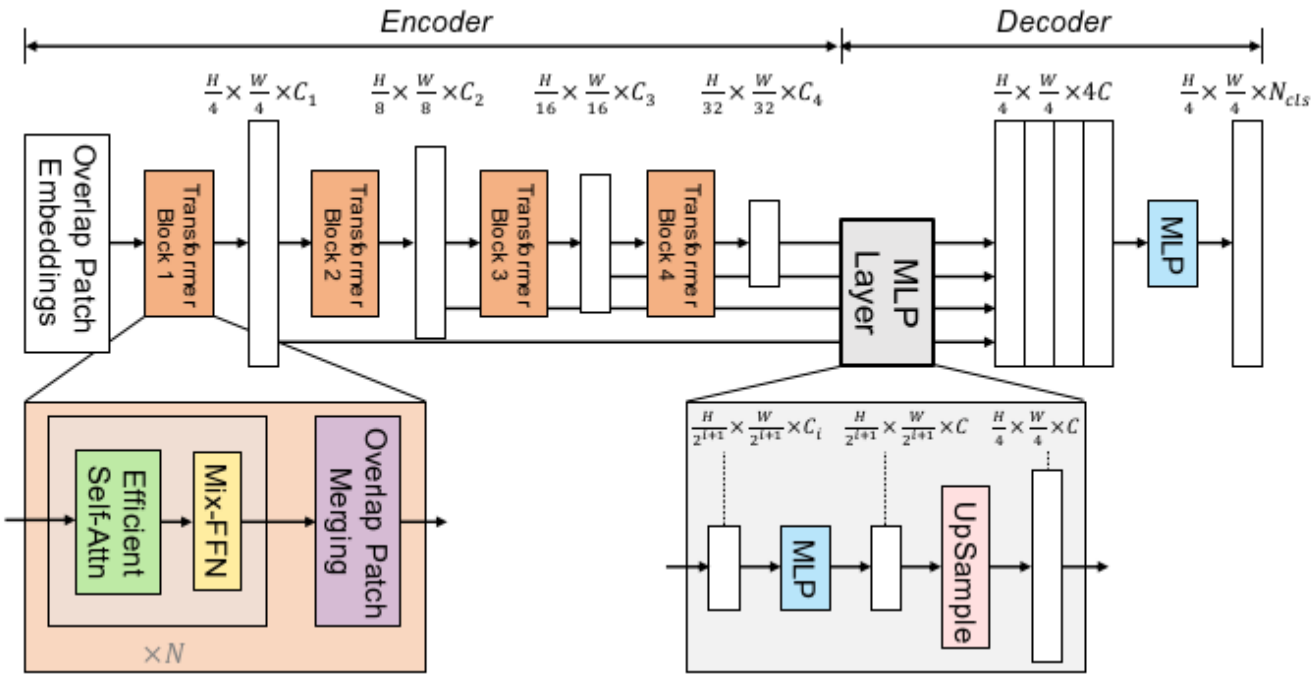

Figure 3 The proposed SegFormer framework consists of two main modules: A hierarchical Transformer encoder to extract coarse and fine features; and a lightweight All-MLP decoder to directly fuse these multi-level features and predict the semantic segmentation mask. “FFN” indicates feed-forward network (Xie et al., 2021).

The validation process:

This process is critical for assessing the model's performance and selecting the best version. A separate dataset, distinct from the training set, is used for evaluation, and metrics like accuracy, precision, recall, and Intersection over Union (IoU) are computed. These metrics provide quantitative measures of the model's ability to classify pixels accurately and delineate fault boundaries. Based on the validation results, the best-performing model is chosen according to specific task requirements.

The iterative training:

The pre-processed data is fed into the SegFormer model, which learns to map input images to pixel-wise segmentation masks. The model's internal parameters are adjusted through backpropagation and optimization algorithms to minimize the difference between predicted and ground truth masks. The training process continues for a specified number of epochs, monitoring metrics like loss and accuracy. The iterative training continues until the model achieves a satisfactory level of fault segmentation performance, as determined by validation metrics and analysis of the model's output.

3. Results and Discussion

We conducted a comprehensive evaluation of SegFormer's performance by examining the effects of different learning rates and backbone models, which are MiT-B1, MiT-B3, and MiT-B5, on the model's performance. Table 1 shows that the Mix Transformer encoder backbone introduced in SegFormer. We utilized various evaluation metrics, such as accuracy and loss value to assess the model's performance. Table 2 presents parameters of each model with vary learning rate in MiT-B1

Figure 4(a) to 4(d) depict the loss values obtained from the trained and validation data using the MiT-B1, MiT-B3, and MiT-B5 backbone models. Lower loss values indicate higher efficiency of the learning model. Among the three backbone models, MiT-B5 demonstrates the lowest loss value, followed by MiT-B3 and MiT-B1, respectively.

Compared with the MiT-B1 results, it is observed that higher learning rates correspond to higher loss values. Additionally, in the case of the MiT-B1 model, the loss values for both the validation and trained data remain parallel across the number of epochs. This indicates that increasing the number of epochs does not significantly impact the loss value for this specific dataset.

In contrast, when training the MiT-B3 and MiT-B5 models for 100 epochs, the stability of their loss values cannot be definitively concluded. However, a higher number of epochs might show a parallel alignment of loss values for both the validation and trained data.

These findings shed light on the relationship between backbone models, learning rates, and loss values. Further analysis and experimentation are recommended to fully understand the stability and convergence behavior of the MiT-B3 and MiT-B5 models with varying numbers of epochs.

Figure 5 is crossline: 1111 which provides a visual representation of the fault recognition results obtained from the experiment. The accuracy of the MiT-b1, MiT-b3, and MiT-b5 models, each trained with a learning rate of 0.001, are 97.6%, 98.5%, and 98.8% respectively.

Figure 6 presents difference between fault interpretation by human and models. Human sometimes leverage personal experience in the interpretation as seeing the white line draw from top of the seismic to the lower part, although the models can capture the discontinuous only where the orange arrows

pointed. It is important to acknowledge that uncertainties exist in seismic data, leading to faults that may go undetected due to limited resolution. The model help capture the fault that has valid evidence then human can utilize its result with experience to identify additional possible fault prediction in the seismic.

Table 1 The Mix Transformer encoder backbone introduced in SegFormer.

Model variant	Depths	Hidden sizes	Decoder hidden size		ImageNet-1k Top 1
MiT-b0	[2, 2, 2, 2]	[32, 64, 160, 256]	256	3.7	70.5
MiT-b1	[2, 2, 2, 2]	[64, 128, 320, 512]	256	14	78.7
MiT-b2	[3, 4, 6, 3]	[64, 128, 320, 512]	768	25.4	81.6
MiT-b3	[3, 4, 18, 3]	[64, 128, 320, 512]	768	45.2	83.1
MiT-b4	[3, 8, 27, 3]	[64, 128, 320, 512]	768	62.6	83.6
MiT-b5	[3, 6, 40, 3]	[64, 128, 320, 512]	768	82	83.8

Table 2 The parameters are set in this experiment in SegFormer.

Backbone Models	epoch	Optimization	Hyperparameter	
			learning rate	batch size
MiT-B1	100	AdamW (PyTorch)	0.001	8
MiT-B1	100	AdamW (PyTorch)	0.01	8
MiT-B3	100	AdamW (PyTorch)	0.001	8
MiT-B5	100	AdamW (PyTorch)	0.001	8

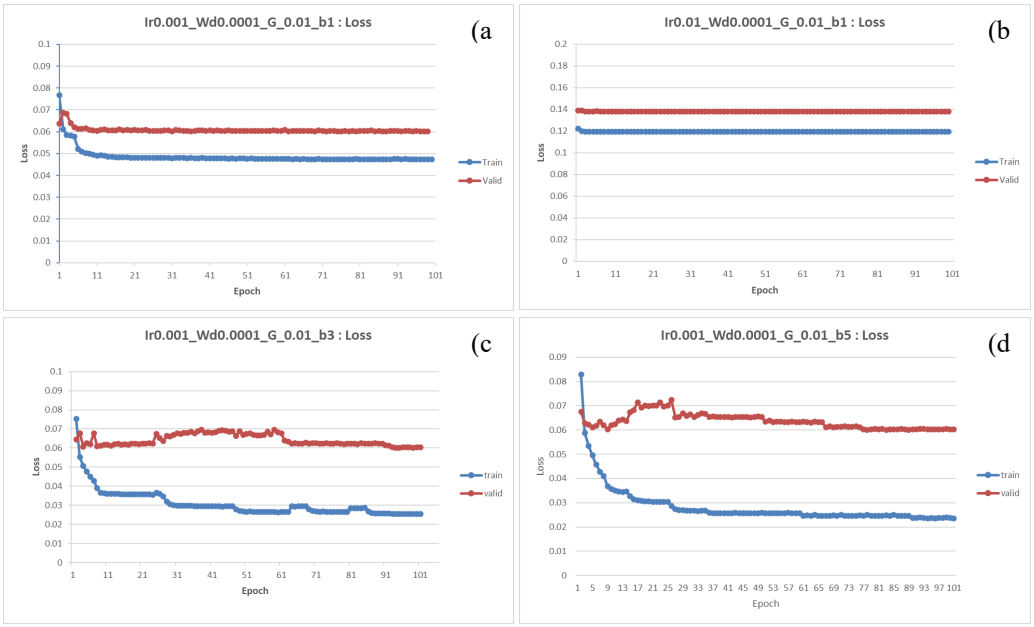

Figure 4 The loss values obtained from each experiment were compared as follows: (a) Experiment with MiT-B1 backbone model and learning rate of 0.001. (b) Experiment with MiT-B1 backbone model and learning rate of 0.01. (c) Experiment with MiT-B3 backbone model and learning rate of 0.001. (d) Experiment with MiT-B5 backbone model and learning rate of 0.001

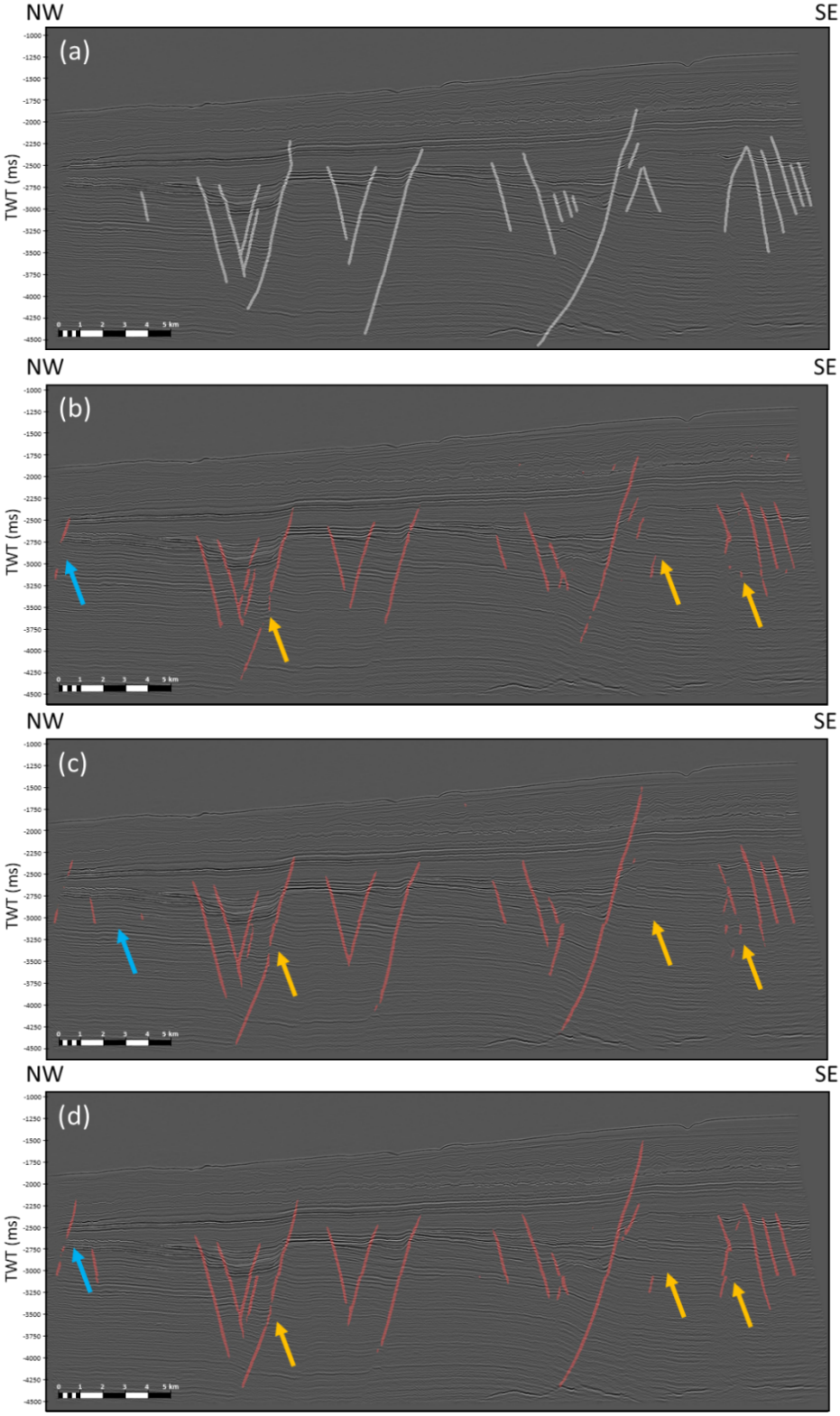

Figure 5 The comparison of fault recognition results is presented in the figure. The subfigures are described as follows: (a) Seismic data and fault interpretation by expert interpreters from the Fault Analysis Group at University College Dublin (An et al., 2021). (b) Seismic data overlayed with fault predictions using MiT-B1. (c) Seismic data overlayed with fault predictions using MiT-B3. (d) Seismic data overlayed with fault predictions using MiT-B5. Orange arrows show the varying outputs of different models and humans' interpretation. Blue arrows indicate the model's interpretation, but humans may not interpret them

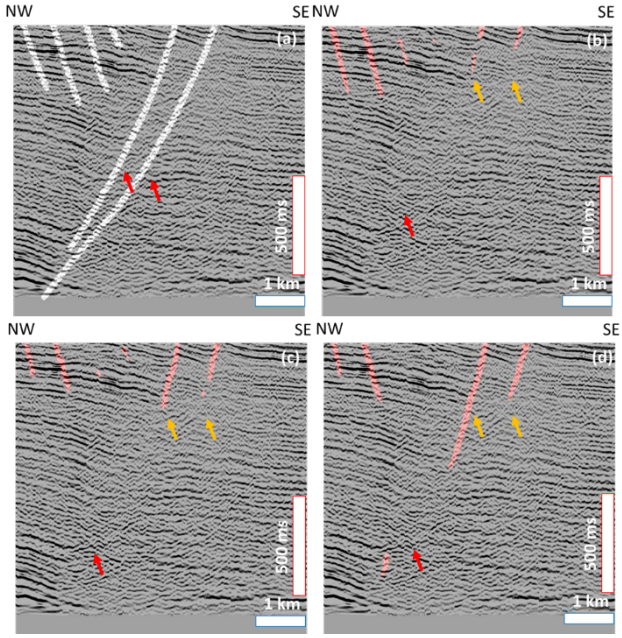

Figure 6 The comparison of fault recognition results is presented in the figure. The subfigures are described as follows: (a) Seismic data and fault interpretation by expert interpreters from the Fault Analysis Group at University College Dublin (An et al., 2021). (b) Seismic data overlayed with fault predictions using MiT-B1. (c) Seismic data overlayed with fault predictions using MiT-B3. (d) Seismic data overlayed with fault predictions using MiT-B5. Orange arrows indicate faults base on discontinuous sequence in seismic section but red arrows point the area that does not show discontinuous sequence in seismic section.

we summarize the impact of the experiments in two main areas:

1.Performance across Backbone Models

In this study, we evaluated SegFormer's performance using different backbone models, specifically MiT-B1, MiT-B3, and MiT-B5. Through extensive evaluation, we observed that SegFormer consistently demonstrates exceptional fault recognition capabilities across all tested backbone models. Notably, the MiT-B5 model achieved the highest accuracy and showed the lowest loss value. However, it is important to note that the MiT-B5 model required the longest training time among the three models.

2.Impact of Learning Rate

We investigated the influence of adjusting the learning rate on the convergence speed and stability of the training process. The results of our study indicate that using a lower learning rate leads to lower loss values, suggesting better convergence. However, it is crucial to note that further experimentation is required to identify the optimal learning rate that balances loss value and convergence speed. Exploring a range of learning rates in future experiments will allow for a more comprehensive understanding of this trade-off.

By providing these insights, our study contributes to the understanding of SegFormer's performance in fault recognition tasks. These findings can guide researchers and practitioners in selecting appropriate backbone models and learning rates, ultimately enhancing the accuracy and efficiency of fault recognition systems based on SegFormer.

4. Conclusion

In conclusion, this paper highlights the superiority of the MiT-B5 backbone model in the context of fault recognition tasks when integrated within the SegFormer architecture. Through a comprehensive evaluation, we have demonstrated the effectiveness of SegFormer in accurately identifying faults in diverse scenarios. The experimental results unequivocally establish MiT-B5 as the best backbone model for fault recognition within the SegFormer framework.

Furthermore, we have investigated the impact of varying learning rates on the performance of fault recognition algorithms integrated within SegFormer. Our findings emphasize the critical role of the learning rate in achieving high accuracy in fault recognition. We have shown that fine-tuning the learning rate is crucial for optimizing the performance of SegFormer for fault recognition tasks.

The data regarding fault recognition systems indicates that such systems can be applied to various geographical areas as long as the training dataset includes different types of faults, such as normal faults, reverse faults, or thrust faults. The study's insights hold practical implications for both practitioners and researchers who are utilizing fault recognition systems based on SegFormer.

The insights gained from this study have practical implications for practitioners and researchers working with fault recognition systems based on SegFormer. By considering the use of the MiT-B5 backbone model and appropriately fine-tuning the learning rate, they can significantly improve the performance of their fault recognition systems. These findings provide valuable guidance for configuring and optimizing SegFormer.

Looking ahead, there are promising research directions to explore regarding the application of SegFormer in fault recognition tasks, particularly in the domain of seismic fault prediction. To further advance the field and expand the practical applications of SegFormer, it is important to investigate its performance on larger and more diverse datasets. By incorporating a wide range of fault recognition scenarios and leveraging these datasets, we can gain a more comprehensive understanding of SegFormer's capabilities and its ability to generalize to various fault recognition scenarios.

This avenue of research holds significant potential for advancing fault recognition techniques and extending the practical utility of SegFormer in the seismic domain. The development of a robust fault prediction model based on SegFormer, informed by a broader evaluation on diverse datasets, would contribute significantly to the field and have implications for various real-world applications.

In summary, this study has made a substantial contribution to the existing knowledge by showcasing the exceptional fault recognition capabilities of SegFormer, particularly when coupled with the MiT-B5 backbone model. We have highlighted the crucial role of the learning rate in enhancing fault recognition accuracy within the SegFormer framework. By providing valuable insights into optimizing learning rates and backbone models, this research aims to advance the field of fault recognition and further enhance the performance of SegFormer-based systems in fault recognition applications.

5. Acknowledgements

Sincere gratitude is extended to PTTEP for providing the first author with a scholarship to study a Master of Science in Petroleum Geoscience at Department of Geology, Chulalongkorn University. Thanks to Australia National Offshore Petroleum Information Management System (NOPIMS) for providing the Thebe dataset.

6. Reference

Alcalde, J., Bond, C.E., Johnson, G., Kloppenburg, A., Ferrer, O., Bell, R. & Ayarza, P. 2019. Fault interpretation in seismic reflection data: an experiment analysing the impact of conceptual model anchoring and vertical exaggeration. *Solid Earth*, 10, 1651-1662.

An, y., Guo, J, Ye, Q, Childs, C., Walsh, J. & Dong, R. 2021. Deep convolutional neural network for automatic fault recognition from 3D seismic datasets. *Computers & Geosciences*, 153 (August 2021), 104776.

Badley, M., Freeman, B., Roberts, A., Thatcher, J., Walsh, J.J., Watterson, J., Yielding & G. 1991. Fault interpretation during seismic interpretation and reservoir evaluation. In: The Integration of Geology, Geophysics, Petrophysics and Petroleum Engineering in Reservoir Delineation, Description and

Management. Proc. Conference. Houston, 1990, AAPG, 224-241.

Biondi, B., 2007. Society of Exploration Geophysicists, European Association of Geoscientists and Engineers, 2007. Concepts and Applications in 3D Seismic Imaging: 2007 Distinguished Instructor Short Course. *Society of Exploration Geophysicists*, 243.

Csillag, F. & Stogicza, Á. 1987. Application of satellite imagery in tectonic interpretation. *Journal of Geodynamics*, 8, 205–219

Cohen, I., Coult, N. & Vassiliou, A.A. 2006. Detection and extraction of fault surfaces in 3D seismic data. *Geophysics*. 71, 21-P27.

Di, H., 2018. Developing a seismic pattern interpretation network (SpiNet) for automated seismic interpretation. *arXiv:1810.08517*.

Di, H., Shafiq, M.A. & AlRegib, G. 2017. Seismic-fault detection based on multiattribute support vector machine analysis. *SEG Technical Program Expanded Abstracts 2017*. Society of Exploration Geophysicists. 2039-2044.

Di, H., Shafiq, M.A., Wang, Z. & AlRegib, G. 2019. Improving seismic fault detection by super-attribute-based classification. *Interpretation*, 7, SE251-SE267.

Fossen, H., 2010. *Structural Geology*. Cambridge University Press.

Gibson, D., Spann, M., Turner, J., Wright, T. & Hale, D. 2012. Fault surfaces and fault throws from 3D seismic images. *SEG Technical Program Expanded Abstracts 2012*, 43, 2094-2102.

Guitton, A. 2018. 3D convolutional neural networks for fault interpretation: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Guitton, A., Wang, H., Trainor-Guitton, W. 2017. Statistical imaging of faults in 3D seismic volumes using a machine learning approach. *SEG Technical Program Expanded Abstracts 2017*. Society of Exploration Geophysicists, 2045-2049.

Guo, B., Li, L. & Luo, Y. 2018. A new method for automatic seismic fault detection using convolutional neural network. *88th Annual International Meeting, SEG, Expanded Abstracts*, 1951-1955.

Huang, L., Dong, X. & Clee, T.E. 2017. A scalable deep learning platform for identifying geologic features from seismic attributes. *The Leading Edge*, 36, 249-256.

Iacopini, D., Butler, R.W., Purves, S., McArdle, N. & De Freslon, N. 2016. Exploring the seismic expression of fault zones in 3D seismic volumes. *Journal of Structural Geology*, 89, 54-73.

Kortström, J., Uski, M. & Tiira, T., 2016. Automatic classification of seismic events within a regional seismograph network. *Computers & Geosciences*, 87, 22-30.

Li, Y., & Vasconcelos, N. 2018. Learning to Count Objects in Natural Images for Visual Question Answering. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4799-4808.

Lisle, R. 2004. Geological Structures and Maps. *A Practical Guide*, 106.

Marfurt, K.J., Kirlin, R.L., Farmer, S.L. & Bahorich, M.S. 1998. 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics*, 63, 1150-1165.

Marfurt, K.J., Sudhaker, V., Gersztenkorn, A., Crawford, K.D. & Nissen, S.E., 1999. Coherency calculations in the presence of structural dip. *Geophysics*, 64, 104-111.

Mohammadpoor, M. & Torabi, F. 2018. Big data analytics in oil and gas industry: An emerging trend. *Petroleum*, 6, 321-328.

Richards, F.L., Richardson, N.J., Bond, C.E. & Cowgill, M. 2015. Interpretational variability of structural traps: implications for exploration risk and volume uncertainty. *Geological Society Special Publication*, 421 , 7-27.

Robein, E. 2010. *Seismic Imaging: A Review of the Techniques, their Principles, Merits and Limitations (EET 4)*. EAGE, 244.

Roberts, A. 2001. Curvature attributes and their application to 3D interpreted horizons. *First Break*, 19, 85-100.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. & Fei-Fei, L. 2014. ImageNet large scale visual recognition challenge. *International journal of computer vision*, 115, 211-252.

Silva, C., Marcolino, C. & Lima, F. 2005. Automatic fault extraction using ant tracking algorithm in the marlim south field, campos basin. *Seg Technical Program Expanded Abstracts*, 24.

Smith, L., Nair, V., & Hinton, G. 2017. A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. 2017. Attention is All You Need. *In Advances in Neural Information Processing Systems (NeurIPS)*.

Wang, Z. & AlRegib, G. 2014. Automatic fault surface detection by using 3D Hough transform. *SEG Technical Program Expanded Abstracts*, 33, 1439-1444.

Wu, X., Liang, L., Shi, Y. & Fomel, S. 2019b. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84, IM35-IM45.

Wu, X. & Fomel, S. 2018. Automatic fault interpretation with optimal surface voting. *Geophysics*, 83, O67-O82.

Xie, E., Wang, W., Yu, Z., Anandkuma, A., Alvarez, J.M. & Luo, P. 2021. *Conference on Neural Information Processing Systems (NeurIPS).*

Yan, Z., Liu, S. & Gu, H. 2019. Fault image enhancement using a forward and backward diffusion method. *Computers & Geosciences*, 131 (June), 1-14.

Yuan, Z., Huang, H., Jiang, Y., Tang, J. & Li, J. 2019. An enhanced fault-detection method based on adaptive spectral decomposition and super-resolution deep learning. *Interpretation*, 7, T713-T725.

Zhao, T. & P. Mukhopadhyay. 2018. A fault-detection workflow using deep learning and image processing. *88th Annual International Meeting, SEG, Expanded Abstracts*, 1966-1970.