Leveraging Visual Prompting to Fine-Tune the Segment Anything Model for Seismic Facies Analysis

Introduction

Seismic interpretation, particularly facies classification, highlights a specific facies object in all seismic data. Current studies show how deep learning architecture such as convolutional neural networks can be used to automatically identify geological objects quickly and with high accuracy. Hall (2016) shows how a classification algorithm called support vector machine is used to identify lithofacies based on borehole measurements. Bilal and Jiang (2022) demonstrated a paleochannel clustering workflow with four state-of-the-art unsupervised methods and used principal component analysis (PCA) to refine and differentiate the results. However, the generalization gap still poses a significant challenge for the data-driven algorithm on different datasets, especially seismic data. To close the generalization gap, highly labelled data is critical. Not only is 3D annotation costly and time-consuming, but there are few resources available for open annotation. Therefore, due to this limitation, models are limited by synthetic datasets and perform poorly when generalized to field datasets, which have a significant generalization gap for various seismic interpretation methods (Jiang and Osypov, 2023). To address these challenges, scientists began searching for a basic model to address the generalization gap.

A major advance in deep learning for image segmentation was achieved by Meta's Segment Anything Model (SAM) (Kirillov et al., 2023). To enable SAM to respond from any given prompt, such as to produce legitimate segmentation mask, such as textual or spatial cues to identify an object, it was created with a viable segmentation task in mind. The advanced architecture consists of a strong image encoder, a prompt encoder, and a lightweight mask decoder, all used in the SAM. This special architecture enables flexible prompting, real-time mask computation and ambiguity detection in segmentation tasks. The original SAM was trained on over 11 million images and provides a broad and rich training data source as it is the largest segmentation dataset to date. Other applications require little rapid engineering as it has exceptional zero-shot performance on a range of segmentation tasks. In addition, it enabled visual prompting during inference time, which provided the interpreters with a good interactive environment to better support people in seismic interpretation and actively guide interpreters to generate results in real time.

In this abstract, we have fine-tuned SAM to create a foundation model for facies classification with multiple visual prompts. We trained the model with multi-facies labels and multi-visual prompts simultaneously. It allows combining the capabilities of manual interpretation and artificial intelligence technology to jointly build an interactive system for seismic interpretation. In our method, interpreter could assign a specific visual prompt to a particular facies class, so our foundation model helps to automatically provide an interpretation based on the fine-tuned SAM.

Methods

Fine-tuning the SAM includes adapting the pre-trained model to the downstream tasks and improving its performance on various seismic interpretation objects, such as facies classification. This process typically requires adjusting the model's weights using a new, annotated dataset, such as PARIHAKA-3D data, that we have implemented in this paper better represents the desired segmentation domain. SAM, as a large-scale vision model for image segmentation, is built on a foundation of transformer-based architectures that provide exceptional capabilities in generalizing a wide range of have shown image segmentation problems. Fine-tuning improves these capabilities by adapting the model's parameters to the downstream tasks.

The fine-tuning process generally includes several important steps: data preparation, model fitting, and training. First, a dataset with annotated segmentation masks is required. We first divided the entire dataset into several small patches and created a dictionary-like database consisting of pairs of seismic data and classified facies. Depending on the specific use case, annotations can range from pixel-wise masks (for semantic segmentation) to object-level boundaries (e.g. segmentation). We considered multiple boxes in a patch as clues to delineate different facies classes (Figure 1). In practice, fine-tuning datasets may require transformations such as resizing, normalization, and expansion to ensure that the model generalizes well under different conditions.

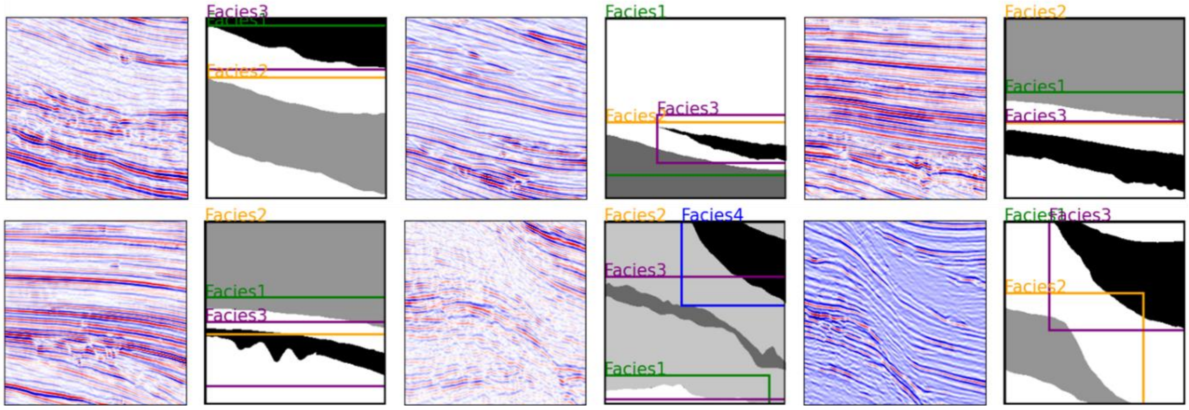

Figure 1: Facies dataset prepared for fine-tuning SAM. In each pair of training data, we created visual prompts with multiple labels (colored boxes), where each visual prompt represents specific facies to fine-tune SAM.

Once the dataset is ready, the next step is to adapt SAM to the new task. This often requires a slight change in the architecture to adapt to the new data. SAM's transformer-based encoder-decoder structure allows flexibility. In this article, we made modifications to the decoder head to output task-specific segmentation maps, such as: B. Facies mapping. In most scenarios, fine-tuning generally focuses on the final layers of the network, where the layers take on the task of generating segmentation masks. These layers are initialized with the pre-trained weights, while the remaining layers are fine-tuned based on the new data. During the fine-tuning step, we freeze weights from the earlier layers of the model because these layers capture general features that are likely to be relevant across different domains. Specifically, we freeze the data encoder and the prompt encoder and optimize only the mask decoder to reduce the computational cost and prevent overfitting (Figure 2).

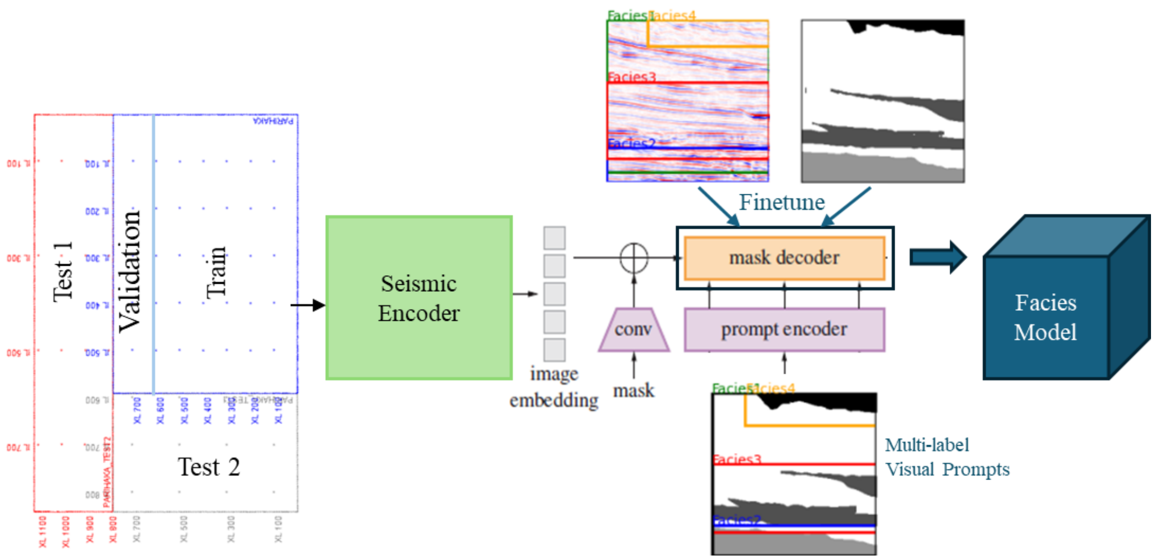

Figure 2: The architecture for fine-tuning SAM using multi-label visual prompts.

During the training process, the model is trained using cross entropy for categorical segmentation tasks. In this case, to prevent overfitting, various regularization techniques such as dropout, early stopping, and learning rate schedules are also applied. The optimizer like Adam is typically used to adjust the model weights, carefully choosing the learning rate to balance fast convergence with stability.

We perform the fine-tuning task on a Nvidia DGX H100 GPU cluster. It is important to periodically evaluate the model against validation data to monitor its performance and ensure that it generalizes well. After the model was fine-tuned, we calculated Intersection Over Union (IoU) for both the fine-tuned model and the original SAM as a base model to compare performance. The comparison between the fine-tuning model and the base model shows significant improvements achieved through the fine-tuning process.

Results

Figure 3 shows a comparison of validation tests between the base model and fine-tuned model seismic facies classification model. During the fine-tuning process, we perform visual prompts with multiple labels to adjust the fine-tuning process. Each prompt or box represents a facies type that is assigned a specific index. This setup provides the opportunity to allow the interpreter to be involved in the classification process by manually assigning specific facies to the training workflow. Therefore, the fine-tuned model will not only predict facies blindly, but also take into account the interpreter's decision to predict the desired facies type. In the validation test, not only are the IoU scores significantly improved by the fine-tuned model, but the results also make more meaningful than the base model in terms of geological interpretation. The mean IoU score over all validation tests improved from 64% for a base model to 88% for the fine-tuned model.

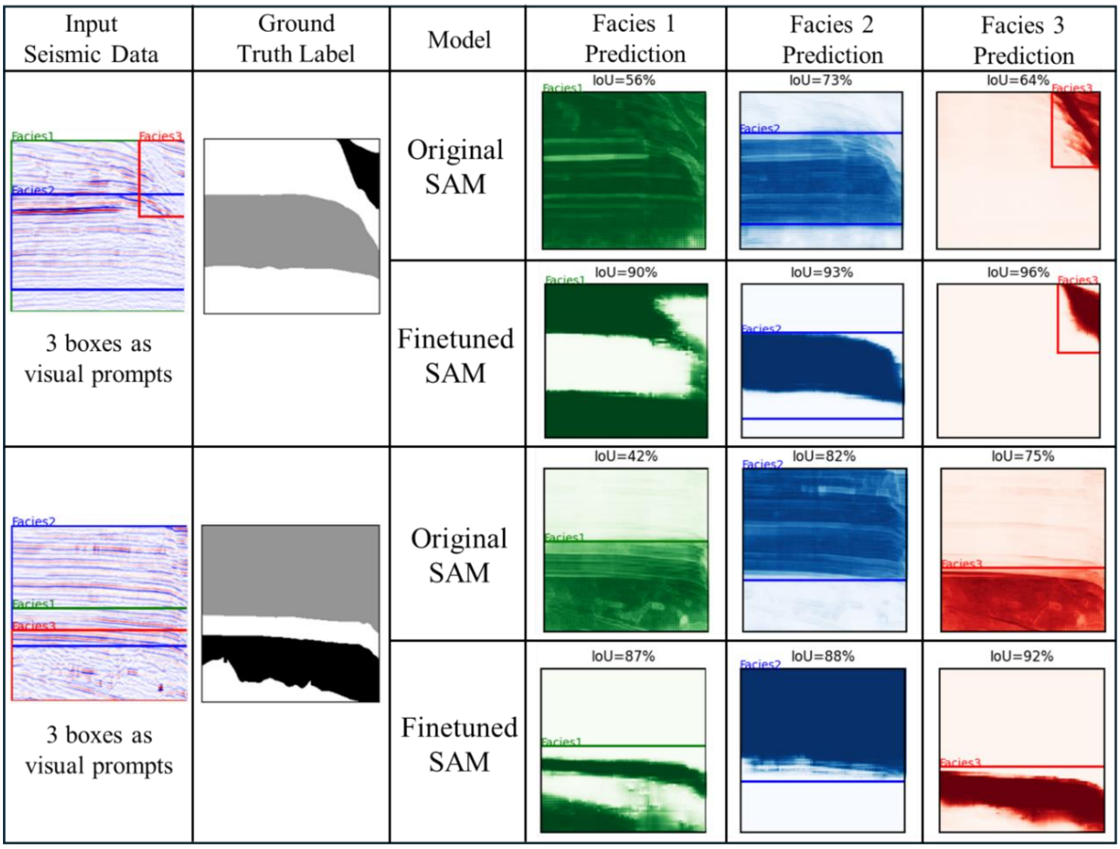

Figure 3: Multiple validation tests of the original SAM and our fine-tuned model. Each color field represents a desired facies type. The IoU scores are significantly improved by fine-tuning SAM with seismic facies data.

We then apply the fine-tuned model to the Test 1 area (Figure 2). Figure 4 shows multi-facies classification with visual prompts with multiple labels to help interpreter distinguish between facies. The final result could also be influenced by the area where the visual prompt is to be applied. This could lead to a similar prediction across two colour patches if the interpreter specifies two facies patches in the same area. In other words, it provides the opportunity to involve the interpreter to leverage their knowledge to achieve high accuracy of trustworthy machine learning results.

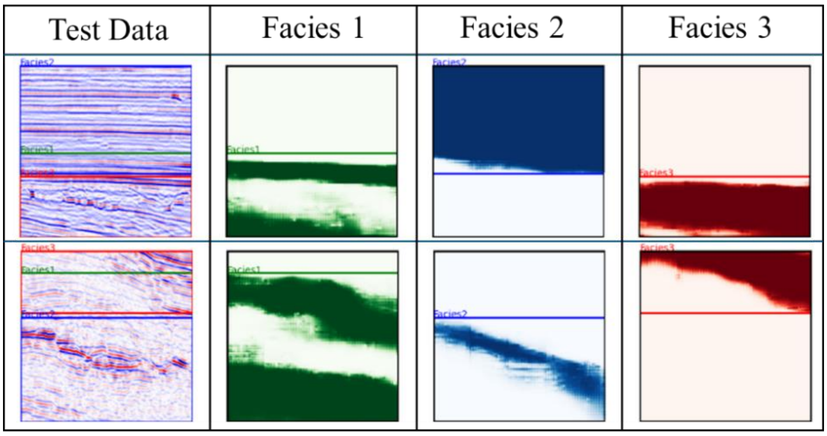

Figure 4: Prediction results with fine-tuned model.

Conclusions

In this abstract, we have fine-tuned a transformer-based segmentation SAM for seismic facies classification. By providing multi-label visual prompts, we validated the ability of a base SAM to produce an accurate classification result when properly fine-tuned, while multi-label visual prompts are analogous to geoscientist's interpretation and provide a practical way to leverage the interpreter's domain knowledge on a machine learning process. Through multiple tests, fine-tuning SAM with downstream task datasets will significantly improve the accuracy and effectively reduce the generalization gap. It offers a promising opportunity to build a generalized model for seismic interpretation.

Acknowledgements

The authors thank the New Zealand Petroleum and Minerals for providing the seismic data used in this research.

References

Bilal, H. and Jiang, F., 2024, Unsupervised Clustering on Paleochannel Detection with Curvature-Based Attribute, IPTC-23429-EA.

Hall, B., 2016, Facies classification using machine learning, The Leading Edge, V35, I10.

Jiang, F. and Osypov, K., Implementation of neural style transfer to mitigate domain discrepancy in the deep learning salt interpretation, 84th EAGE Annual Meeting.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W., Dollar, P. and Girshick, R., 2023, Segment Anything, arXiv:2304.02643.