Submission code: 90BBPJY90G

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Enhancing salt interpretation with self-supervised learning: a study using DINO

Daniel Miranda (Petrobras), Alan Albano (Petrobras), Vitor Batista (Petrobras)

Enhancing salt interpretation with self-supervised learning: a study using DINO

Please, do not insert author names in your submission PDF file

Abstract Summary

This study explores the use of self-supervised learning (SSL) to train deep neural networks without labeled data, focusing on convolutional networks rather than the Vision Transformers commonly used in recent seismic data studies using SSL. The authors pretrain a ResNet-101 model using the DINO SSL method and 57 seismic volumes from offshore locations in Brazil. The pretrained ResNet-101 is then integrated into the DeepLabV3+ architecture as an encoder and evaluated for the task of salt interpretation. The model's performance is compared to a baseline pretrained on the supervised ImageNet-21K dataset. The results demonstrate that SSL pretraining with seismic data improves interpretation quality, highlighting the potential of SSL approaches for seismic analysis tasks.

Introduction

Self-supervised learning (SSL) is a machine learning paradigm that enables deep neural network (DNN) training without labeled data by leveraging extensive unlabeled datasets to build robust representations. These DNNs establish a foundation model for diverse downstream tasks. Sheng et al. (2025) implemented the Masked Autoencoder (MAE) SSL framework to train Vision Transformers (ViTs), demonstrating that pretrained ViTs achieve state-of-the-art performance in classification, segmentation, and interpolation tasks. Sansal et al. (2025) used MAE to 3D ViT pretraining, proving their effectiveness as encoders in salt segmentation DNNs with zero-shot inference capabilities.

In this work, we use the DINO SSL method, developed by Caron et al. (2021), to pretrain a ResNet-101 model on depth-migrated final-stack seismic data. The pretrained network is subsequently integrated as an encoder into the DeepLabV3+ architecture for quantitative evaluation of salt interpretation performance. We benchmark results against a ResNet-101 baseline pretrained on the supervised ImageNet-21K (IN-21K) classification benchmark. Both architectures were trained on two interpreted seismic lines and rigorously evaluated across 40 interpreted lines.

Methodology

The methodology in this work is divided into three main steps: preparation of the pre-training dataset, pre-training using the DINO method, and salt segmentation. Each of these steps is described below.

For the preparation of the pre-training dataset, we used 57 seismic volumes from offshore locations in Brazil, all of which are depth-migrated final stacks. From these volumes, we extracted several patches of size 512x512 from the inline and crossline directions of these seismic volumes.

In the pre-training phase, we employed the DINO method to train a ResNet-101 model on the prepared dataset. The model was trained using an image size of 256x256.

For the final step, we trained a DeepLabV3+ model with a ResNet-101 as the feature extractor for the semantic segmentation of salt. The training dataset consisted of one inline and one crossline, and the model underwent full fine-tuning. Two model versions were trained: one with the ResNet-101 pre-trained using the DINO method (Deeplab-DINO) and another with the ResNet-101 pre-trained on the supervised ImageNet-21K task (Deeplab-IN21K). After training, inference was performed on an entire seismic depth-migrated volume. The volume dimensions are 1430 × 2567 × 2400, where 1430 and 2567 correspond to the inline and crossline directions, respectively.

Results

We compared the salt segmentation performance of the Deeplab-Dino and Deeplab-IN21K models by calculating the mean Intersection over Union (mIoU) across 40 interpreted lines. Figure 1 illustrates the distribution of these lines, and the quantitative results are presented in Table 1.

The mIoU values for both models are very similar, with Deeplab-Dino achieving a slightly better score. A qualitative comparison of the top of the salt body, shown in Figure 2, highlights the differences between the models. Deeplab-Dino produces a more continuous and geologically coherent top-of-salt interpretation. These results confirm that the pretraining with the DINO method contributes positively to the model's performance.

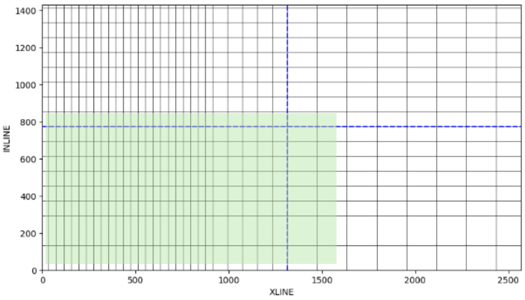

Figure 12: Distribution of interpreted lines. The training lines are displayed in blue. The green area is where the top of salt is displayed in Figure 2.

Model	Deeplab-Dino	Deeplab-IN21K
mIoU	**0.92**	0.91

Table 1: mIoU on teste lines.

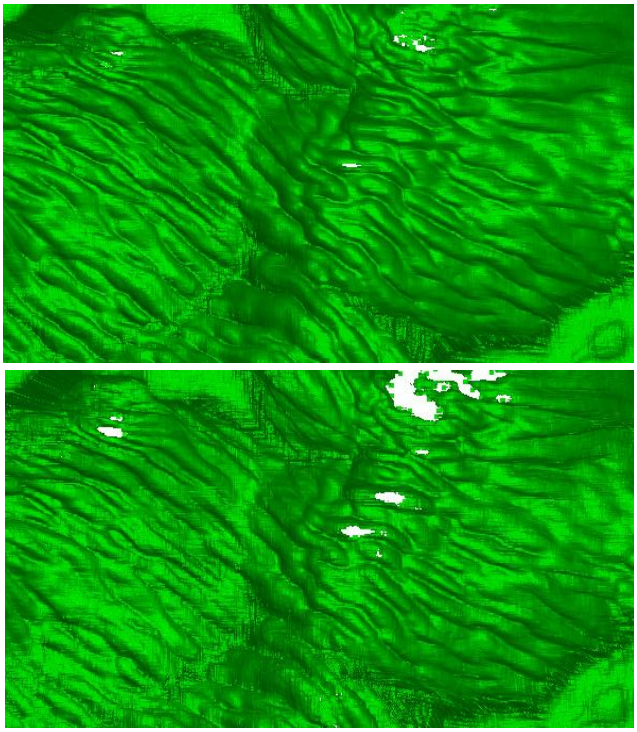

Figure 21: Top of salt body. On the top image, we show the result of Deeplab-Dino. On the bottom, the result of Deeplab-IN21K.

Conclusions

This study investigates the application of the DINO self-supervised learning method for pretraining ResNet-101 models using seismic data and evaluates its effectiveness in the salt interpretation task. The comparison between Deeplab-DINO and Deeplab-IN21K demonstrated comparable mean Intersection over Union (mIoU) metrics. However, Deeplab-DINO produced salt body geometries that were more geologically coherent, indicating that SSL pretraining contributes positively to interpretation quality. While the experiments focused on full fine-tuning of the models, future work exploring frozen layers may provide additional insights into the problem and the broader applicability of SSL techniques in seismic data analysis.

References

Sheng, H., Wu, X., Si, X., Li, J., Zhang, S. and Duan, X., 2025. Seismic foundation model: A next generation deep-learning model in geophysics. Geophysics, 90(2), pp.IM59-IM79.

Sansal, A., Lasscock, B. and Valenciano, A., 2025. Scaling Seismic Foundation Models. First Break, 43(2), pp.69-75.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P. and Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9650-9660).