Submission code: 8PN57VBKLY

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

From space to time: interpolating seismic images using video- and image-based diffusion models

Pedro Gil Couto (PETROBRAS - Petróleo Brasileiro S.A.), Thales Mesentier (PETROBRAS - Petróleo Brasileiro S.A.), Vitor Giudice (PETROBRAS - Petróleo Brasileiro S.A.), Eduardo Bucker (SPASSU), Joziani Vieira (SPASSU)

From space to time: interpolating seismic images using video- and image-based diffusion models

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

Abstract Summary

Seismic imaging is fundamental to Oil & Gas activities, particularly to exploration and reservoir modeling. Therefore, increasing quality and fidelity, and properly and quickly extracting insights is paramount to expedite industry workflows. Considering machine learning techniques, many different approaches have been used. However, when dealing with seismic imaging in post-stack seismic volumes, most approaches treat image slices either as traditional machine learning training data – randomly shuffling inputs, not considering the strong spatial correlation that subsequent seismic slices have in a seismic volume – or resort to computationally expensive three-dimensional convolutional approaches. This work covers an initial attempt at a paradigm shift, proposing interpreting the third spatial dimension as if it was time, aiming at using seismic imaging spatial correlation information during inference while avoiding the high costs of traditional three-dimensional processing.

Introduction

Seismic imaging is fundamental to Oil & Gas activities, particularly to exploration and to reservoir modeling. Therefore, numerous approaches have been tried and used both to enhance data quality and to extract – or at least facilitate the extraction of – insights from the available data.

When considering machine learning approaches, inference is often conducted on an image-based level – composed of 2D seismic slices. To simplify development, most machine learning techniques treat seismic image slices as independent and identically distributed (IID) data, randomly shuffling input data. Considering the highly correlate nature of adjacent slices extracted from a seismic volume – and how important correlation analysis is to traditional, human-based, insight extraction activities – this approach potentially discards significant information, while ignoring important characteristics of the data. On the other hand, examples of machine learning methods that try to take spatial correlation into account for inference usually rely on very computationally expensive methods, such as three-dimensional convolutions.

Finding an efficient solution for this standoff is an important development for machine learning-based seismic imaging processing. In this work, we suggest an initial attempt at addressing this issue by conducting a paradigm shift: instead of treating seismic volumes as three-dimensional from a purely spatial perspective, we propose dealing with the third dimension from a time perspective. That is, we propose treating seismic slices as images, and seismic volumes as images changing through time, such as frames in a video. This allows us to apply consolidated video-based methods to seismic volumes, adding new experimental possibilities for multiple workflows.

To start testing this approach, we choose one of our current challenges. In a couple of different exploratory scenarios, the entirety of an area of interest is covered by seismic acquisitions, but not

all of it by 3D acquisitions. In such areas, where a mix of 2D and 3D acquisitions exists, being able to merge all available data into a single, complete, 3D volume could yield valuable information for interpretation. To generate the necessary 3D volumes, however, it is necessary to interpolate the 2D data with high fidelity and consistency, using the available 3D data as training material for the interpolation model.

In the following sections, we describe the tackling of the interpolation challenge, evaluating the benefits and shortfalls of using video-based techniques in relation to other approaches.

Method

We develop two different interpolation models to compare results. The first model (Wang and Golland, 2023) – henceforth named *traditional model* – uses conditioned diffusion-based image generation to interpolate between two arbitrary 2D seismic slices. The second model (Danier et al., 2024) – henceforth named *video-based model* – employs a diffusion-based video interpolation method to generate interpolated images between two arbitrary seismic slices. For both models, as our domain is significantly different from the original training data, we conduct fine-tuning.

To generate training data, we select 3D seismic volumes and extract vertical slices from them in both acquisition directions. Then, we *decimate* slices according to various interpolation distances. For both models, the training data consists of image triplets: two source images to be interpolated, and the ground truth image, corresponding to the equidistant middle slice. For simplicity, the distance between the middle slice and each source image is named *stride*.

First, training sets for single, arbitrary strides are created (e.g. 5, 10, 20). In the acquisitions we used, the geographic distance between slices was of 25 meters. Therefore, a stride of 20 – amounting to a full distance of 40 slices between source slices – represents a distance of 1 kilometer. Then, generic sets are also built, composed by many different strides. The same procedures are conducted to generate test sets. Training and test sets are extracted from separate, distant areas of each volume – to maximize structural differences between slices contained in the training and test sets, minimizing indirect data leakage.

To train the models, we use a parallelized architecture, leveraging Slurm (Yoo et al., 2003) and Pytorch's Distributed Data Parallel (Li et al., 2020) to distribute multi-GPU and multi-node training. For each model and training set pair, 2 of our cluster nodes are used. Each node consists of 256 vCPU cores (from two AMD EPYC 7742 processor), 2 TB of RAM and 8 NVIDIA A100 80GB GPUs.

A plethora of different training procedures are undertaken, considering different dataset sizes and distributions and different epoch amounts, but for the sake of conciseness, only the final setups are presented in this work. For the traditional model, data requirement is extensive, so final training procedures take 1000 epochs and last over a full day to complete. The video-based model, however, requires much less repetition, being successfully convergent in less than 100 epochs and totaling no more than a couple hours of total fine-tuning time.

To infer result quality, we generate visualization examples and submit them to analysis by geosciences specialists. As with most O&G applications – and with most computer vision tasks – calculated metrics are usually incapable of fully evaluating and explaining results. Hence our focus in the generation of result visualizations.

Results

Generated visualizations for the traditional model can be seen in Figures 1 and 2. They are generated by interpolating test set slices with varied strides. Obtained results are promising and, according to the inquired experts, the model shows great capability of retaining general image structure and

low-frequency tendencies across different strides. However, it is possible to detect performance deterioration as distances increase to very large amounts.

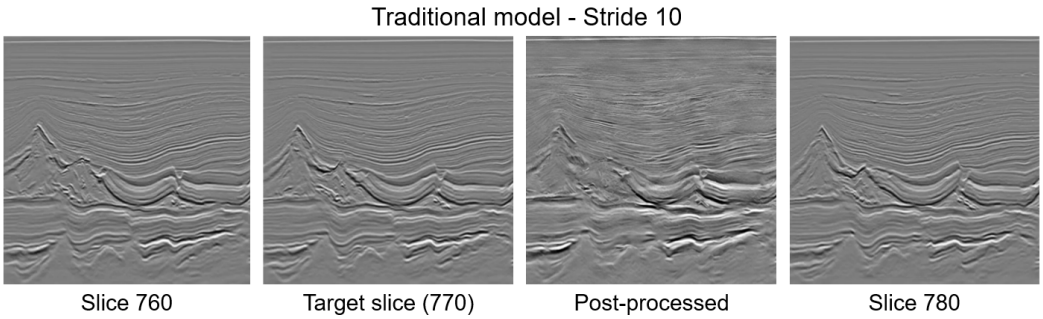

Figure 1: Visualization for the traditional model: stride 10 (500 meters).

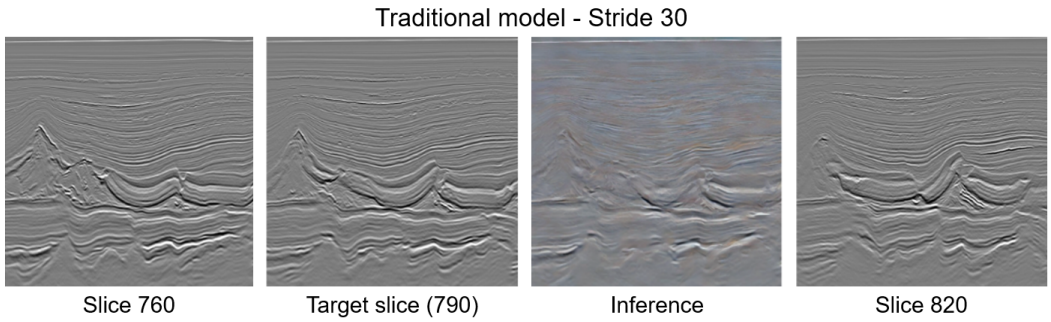

Figure 2: Visualization for the traditional model: stride 30 (1500 meters).

Also, the model has a tendency to input extra, unnecessary channels into the seismic image, generating unwanted color noise and shortening the image's dynamic range. To deal with this behavior, a post processing pipeline is implemented, and is shown in Figure 1 to present comparison to the original outputs.

For the video-based model, visual results can be observed in Figures 3 and 4. They are generated in the same way as for the traditional model. This model excels at smaller strides, but also has deteriorating performance with increasing strides – tending towards an average of the source images. This behavior is expected from a video interpolation approach, as, in general, videos show restricted variations between subsequent frames. The interpolation method, taking this into account, uses, among other things, the calculated difference between the source images as part of the diffusion conditioning. Therefore, as we reach major differences between the images to be interpolated, the model reaches a performance limit. Even so, results are promising, especially considering that the model consumes significantly less time, resources, and data to be trained.

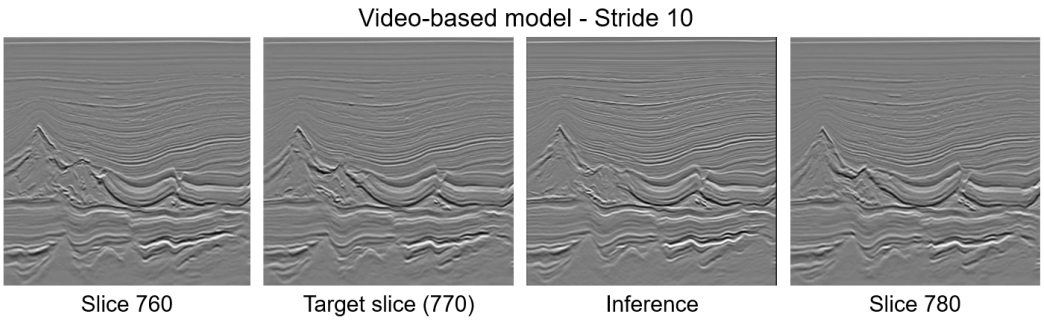

Figure 3: Visualization for the video-based model: stride 10 (500 meters).

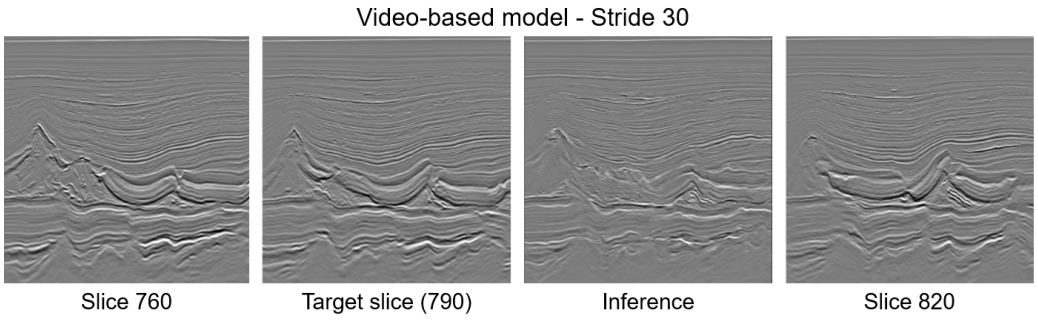

Figure 4: Visualization for the video-based model: stride 30 (1500 meters).

Conclusions

In this work, we presented an initial attempt at a paradigm shift in seismic processing, interpreting the third spatial dimension of seismic volumes as if it was time. We used a seismic interpolation challenge to illustrate the behavior of such approach, and compare it to traditional methods. Presented results are promising, although preliminary, and offer opportunities for further investigating the proposed methodology. Future works involve expanding the downstream workflows experimented with, using the new paradigm for object detection and segmentation tasks, and deepening and standardizing comparison methodologies, undertaking profiling tests and metrics-based performance comparisons.

References

Danier, D., F. Zhang, and D. Bull, 2024, Ldmvfi: Video frame interpolation with latent diffusion models: Proceedings of the AAAI Conference on Artificial Intelligence, **38**, 1472–1480.

Li, S., Y. Zhao, R. Varma, O. Salpekar, P. Noordhuis, T. Li, A. Paszke, J. Smith, B. Vaughan, P. Damania, and S. Chintala, 2020, Pytorch distributed: Experiences on accelerating data parallel training: CoRR, **abs/2006.15704**.

Wang, C. J., and P. Golland, 2023, Interpolating between images with diffusion models.

Yoo, A. B., M. A. Jette, and M. Grondona, 2003, Slurm: Simple linux utility for resource management: Job Scheduling Strategies for Parallel Processing, Springer Berlin Heidelberg, 44–60.