CAMBRIDGE
UNIVERSITY PRESS

TRANSLATIONAL ARTICLE

Data-centric strategies for deep-learning accelerated salt interpretation

Apurva Gala [ORCID] and Pandu Devarakota

Shell Information Technology International Inc., Shell Technology Center Houston, TX, USA
Corresponding author: Apurva Gala; Email: apurva.gala@shell.com

Abstract

Deep learning (DL) has become the most effective machine learning solution for addressing and accelerating complex problems in various fields, from computer vision and natural language processing to many more. Training well-generalized DL models requires large amounts of data which allows the model to learn the complexity of the task it is being trained to perform. Consequently, performance optimization of the deep-learning models is concentrated on complex architectures with a large number of tunable model parameters, in other words, model-centric techniques. To enable training such large models, significant effort has also gone into high-performance computing and big-data handling. However, adapting DL to tackle specialized domain-related data and problems in real-world settings presents unique challenges that model-centric techniques do not suffice to optimize. In this paper, we tackle the problem of developing DL models for seismic imaging using complex seismic data. We specifically address developing and deploying DL models for salt interpretation using seismic images. Most importantly, we discuss how looking beyond model-centric and leveraging data-centric strategies for optimization of DL model performance was crucial to significantly improve salt interpretation. This technique was also key in developing production quality, robust and generalized models.

Impact Statement

Adapting and developing effective deep-learning models for real-world applications and specialized data domains is one of the standing challenges for sustainable deployment of deep-learning models. In this paper, we discuss the challenges in adapting DL models for salt interpretation, a computer vision task using seismic data. Data-centric AI is an emerging field that focuses on developing data engineering pipelines that allow high-quality training data generation to optimize performance of DL models. Our work leveraged data-centric techniques to adapt and optimize DL models to perform salt interpretation robustly and was crucial in training and deploying these models in real-world solutions.

---

1. Introduction

A considerable amount of research in deep learning (DL) and AI modeling has been focused on advancing the algorithms: DL architectures, hyper-parameter tuning, and so on, in other words, model-centric performance optimization. Prototyping a DL model for a particular task, for instance, a typical computer vision task like semantic segmentation of natural images, can be quickly accomplished. Using large,

curated benchmark datasets and pretrained backbones, popular neural network architectures can achieve very high segmentation accuracy.

A multitude of data augmentation techniques can be further leveraged to mitigate lack of variability in natural image datasets (Cubuk et al., 2020; DeVries and Taylor, 2017). Synthetic data generation and self-supervision techniques are also widely applied to address data scarcity or lack of annotated data.

However, training robust and reliable computer vision DL models to specialized application domains like seismic interpretation poses quite a unique data problem. Seismic images are unlike natural images as they are not acquired from visual sensors, however, are predominantly used to perform visual tasks that are perceptually driven, like salt interpretation.

Salt interpretation using seismic images is a crucial step for seismic imaging and subsurface velocity model building workflow. Salt is one of the most challenging geological layers to interpret, as it "moves" across the subsurface in response to stresses and strains from other geological layers. The goal of seismic imaging is to improve the quality of the seismic image, typically under the salt to reduce the imaging uncertainty and accurate localization of geological layers to aid well positioning. Accurate delineation of salt from seismic images leads to a better understanding of subsurface geology, which, in turn, leads to better quality seismic images. Thus, salt interpretation and improving seismic image quality represent the classic causality dilemma. Salt interpretation is exceedingly time and effort intensive and leveraging DL to accelerate this step can result in faster turnaround times for seismic imaging workflows and considerable cost savings.

The goal of the effort described in this paper was to automate salt interpretation using DL models, so that DL models can interpret 80–90% of the salt and the domain experts can more efficiently use their time and expertise in interpreting salt in challenging geology or poorly imaged areas (Alkan et al., 2022). Hence, the requirements for the DL models were (1) detect salt with sufficient accuracy and (2) integrate into an existing salt interpretation workflow tool to enable domain experts to leverage the model predictions to complete their interpretations. The second requirement is production-specific and requires the model to be robust and produce consistent results under unknown production conditions.

With seismic data, perceptually meaningful descriptors like color, texture, and so on are no longer applicable. Seismic processing algorithms like noise removal or enhancement filters are focused on boosting the frequency fidelity of the signal as opposed to perceptual fidelity which introduces large variations in data. Traditional computer-vision DL models are tuned to overcome variations in illumination, sensor angle, object pose, and sensor inadequacies, which tend to be less relevant to seismic data. Hence, taking these fundamental and domain-specific distinctions in type and nature of data into modeling considerations is critical to effectively training DL models for domain-specific tasks. In addition, training production-grade DL models deployed as a part of digital software adds further complexity to the generalization and robustness of models.

In this paper, we present our work on training DL models for salt interpretation using seismic images. We will describe the challenges in adapting semantic segmentation using DL to the domain of seismic interpretation and their subsequent deployment. We will present the data-centric strategies that turned out to be the key factor in training production-quality and robust DL models. More specifically, the impact of quality of training data and labels (required for supervised learning) had on DL model training and performance and how mitigating these factors through the data engineering pipelines led to stable model training and re-training. In addition, we will also discuss a data-conditioning metric which allows to predict the acceptability of segmentation performance generated by DL model inference, given a particular test dataset.

2. Domain data challenge

Velocity model building provides interpretative input to improve seismic imaging quality for the resolution of subsurface depth structures needed for accurate well positioning for oil and gas production. Salt interpretation plays a critical role in velocity model building for imaging complex salt
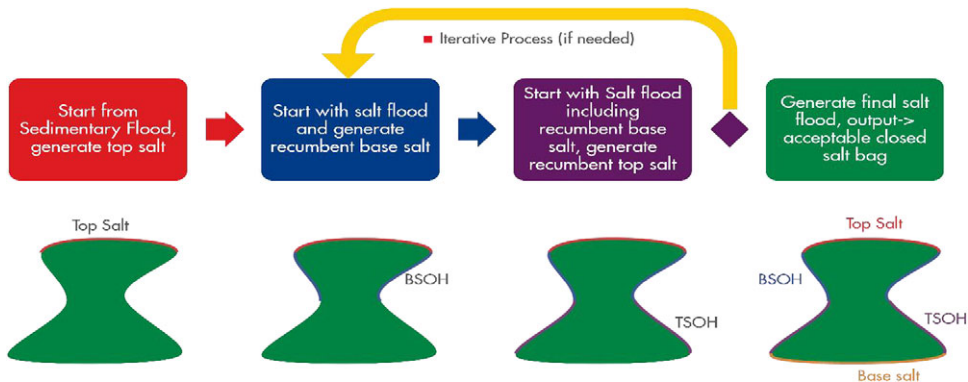
Figure 1. *Top-down salt interpretation (iterative) workflow.*

formations in subsalt and pre-salt rich prospects such as the Gulf of Mexico (GoM), Brazil, and the North Sea.

Typically, salt interpretation, as shown in Figure 1, is a top-down, iterative process, where the interpreter starts with an "incomplete" seismic image (referred to as sedimentary flood image as the initial velocity model only consists of sedimentary structures and no salt in it), and first interprets or picks the top-edge (top salt) of the salt body. Then, using this salt edge location, the seismic image is then re-generated (migrated to generate the so-called salt flood image), with the intent that this image will now better highlight the deeper regions of the salt body, making detecting its deeper edges more effective. The velocity model is then updated with the interpreted salt edges, and the migrated seismic image is subsequently used to pick the deeper edges, and this process continues till the bottom-edge (base salt) of the salt body is interpreted and the entire salt body (salt bag) is well-defined and integrated into the velocity model. Thus, intermediate salt interpretation steps are done using seismic images whose quality with respect to depth is still improving; in other words, seismic image quality is far from optimal.

Furthermore, due to the limitation in the acquisition geometry, the salt edge detection or interpretation is extremely challenging when detecting flanks or overhang interfaces of the salt: base-salt overhang (BSOH) and top-salt overhang (TSOH) which is required to completely describe the 3D object for velocity model building for production processing. Salt flanks and overhangs are typically poorly illuminated due to null space, showing little to no edge in the seismic image. In addition, the existence of overhangs is dependent on salt geometry, and not every salt interpretation will have an overhang. Consequently, the number of surveys with reliably interpreted overhangs are few and far between; hence, less data exist to train the overhang detection model. This process is expectedly time consuming, many times requiring significant human intervention depending on the complexity of salt geometry to be interpreted.

In this effort, we trained distinct DL models for top-salt, overhangs, and base-salt to augment the existing salt interpretation workflow in its various stages by providing the interpreter with interpretation starting points at each stage in the process.

Developing DL models that are robust and well-generalized that produce "good enough" salt interpretation results, however, presents multiple challenges from domain complexity to data quality and more (Shi et al., 2019; Alkan et al., 2022). Seismic datasets are 3D and typically cover large geographical areas. Figure 2 shows a 2D seismic image (on the left), a cross section of a 3D seismic volume, and the same image with the salt edges depicted (on the right), and the salt edges highlighted with yellow and red lines are referred to as top-salt and base-salt interpretations, respectively. Different datasets go through different acquisition and processing techniques, hence vary widely in signal-to-noise ratio and frequency content. In addition, different geographical locations present varying geology and complex geometries of the subsurface structures. The quality of seismic images is determined by coherence and continuity of the seismic reflectors. The quality gap between legacy datasets and newly acquired datasets is quite significant, which makes collecting a large corpus of high-quality, representative data difficult.
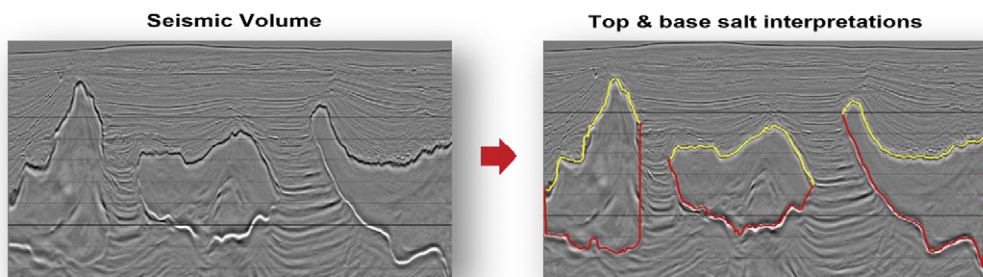
Figure 2. *The figure shows a 2D seismic slice (one the left) and the top (top-edge) and base (bottom-edge) salt interpretations overlaid on seismic to highlight the edge detection task (on the right).*

Salt interpretation is both subjective and challenging when detecting uncertain edges, overhangs, and base salt interfaces as seismic in deeper areas typically suffers from poor illumination, showing little to no edge delineation, inconsistent loops, and poor signal to noise. In such cases, interpreters rely on other data sources, their prior knowledge of the survey area, and their years of experience to imagine and define salt delineations.

However, the domain expertise and experience of interpreters is not well captured in any tangible data source, and hence the DL models rely only on the processed seismic image to detect salt edges. Thus, detecting salt flanks or overhangs in poorly imaged areas remains as a subjective and challenging problem for the DL model, as the input data often do not contain strong visual signatures of salt edges.

In addition, in practice, when interpreters rely on their domain knowledge and experience, it results in multiple interpretations (or labels) for the same seismic survey and significant deviations and inconsistencies in manual picks. Hence, labels are far from producing the so-called "ground truth," making it harder to find and extract good labels for supervised learning (Gala et al., 2022). Furthermore, the training labels are themselves interpretations hence subjective and biased.

The seismic data challenges for salt interpretation can be summarized as follows:

1. *Suboptimal seismic data quality*: The seismic image is iteratively improved through the salt interpretation workflow, and the seismic used during each step of salt interpretation is always suboptimal or not the final "acceptable" image quality.
2. *High degree of freedom in representative data*: High variability in geological and geometric complexity of salt necessitates that the training data be diverse and large enough to train a well-generalized DL model. However, the quality of datasets also varies widely and hence is challenging to collect high-quality representative datasets to train robust models.
3. *Challenging deeper salt edges*: Detection of salt flanks or overhangs is furthermore challenging, as the data around salt flanks are typically poorly imaged and hence difficult to train a DL model. In addition, the amount of data available to train DL models to detect overhangs is limited, as these salt flank structures are not always present in all the places as compared to top-salt or base-salt.
4. *Imperfect labels*: Label themselves are interpretations and in poorly imaged areas can vary and hence are imperfect at best.

3. Challenges in salt interpretation DL model development

The previous section described the salt interpretation workflow, the goal of developing DL models for salt interpretation, and seismic data imperfections and the complexity it poses to training DL models. In this section, we will discuss the approach we took to training generalized DL models.

Salt interpretation was posed as a typical computer vision edge detection problem, and a typical model-centric development workflow was adopted: (1) collect large corpus of representative training data, (2) pre-process data and create the training data artifacts, (3) train simple architectures like CNN-based U-nets and hyper parameter tunning to optimize performance, and (4) experiment with more complex DL architectures to address domain-specific challenges.

3.1. Collect large corpora of representative training data

We had access to several seismic data volumes (each survey spanned across hundreds of kilometers) and salt interpretations, acquired over long periods of time (legacy data) and from many different geographical areas. These are used to create a large corpus of training pairs consisting of seismic and labels derived from existing salt interpretations. To accommodate large variability in acquisition and processing techniques, the training data were selected across these variables so that the trained models are robust to these variabilities.

Subsurface geology and salt geometries vary depending on geographical regions. In addition, salt is interpreted differently depending on the underlying local geology, so labels used for different regions are also distinct, in terms of their alignment with the underlying seismic signal. Hence, different DL models were trained for different geographical regions using training data from the same geographical regions. For instance, training a GoM specific DL model was trained using data from only GoM assets encompassing different geologies and subsurface geometries that exist within GoM. This GoM model was then deployed for seismic interpretation throughout all the assets across GoM.

3.2. Create training data from 3D seismic volumes

Seismic data typically has quite a large dynamic range of amplitudes. The amplitude distribution resembles a normal distribution with long tails, as shown in Figure 3 (image on the right). The salt edges coincide with either positive (shallower salt edges) or negative (deeper salt edges) peaks in amplitudes. In addition, depending on the type of acquisition and processing, the datasets go through the dynamic ranges across datasets vary widely as well. Numerous data normalization techniques were implemented, including log scaling, z-score scaling, and scaling to range (-1 to $+1$). Empirically, we found that simple scaling to range (-1 to $+1$) worked best, as it preserved the complete dynamic range of each dataset which was essential to retain the discrimination between dominant edges and other edges in the data.

The goal of a salt interpretation DL model was to learn the mapping between seismic and salt edge. Seismic data are typically a 3D volume with the x and y axes representing the lateral cross sections and the z-axis representing the depth. In this effort, to align the DL models with top-down salt interpretation
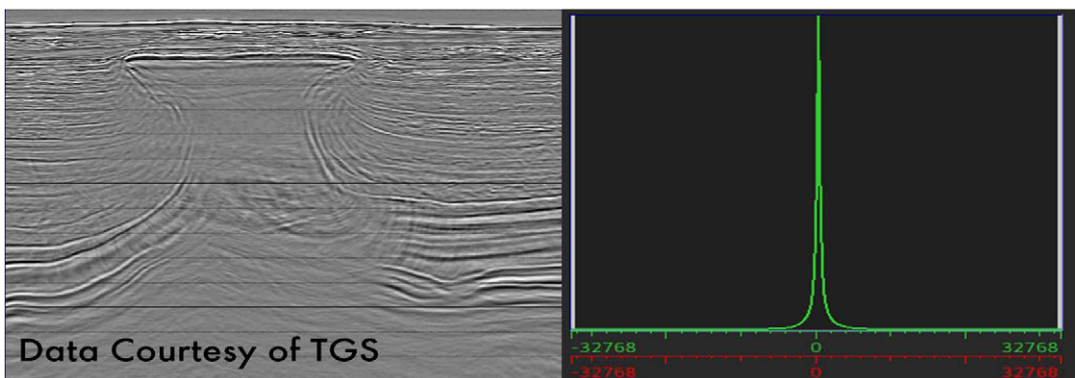

Figure 3. *2D seismic slice from a 3D seismic dataset and the distribution of the seismic amplitudes (wide dynamic range).*

workflow, a predominantly 2D strategy was adopted, where 2D "tiles" were extracted from the lateral slices. Depth slices were not included in the training data. A large dataset of 2D tiles of predetermined size was extracted from cross sections of seismic volumes and the corresponding labeled volume.

3.3. Training generalized salt interpretation DL models

Salt interpretation models were designed to identify and delineate the salt-sediment boundary effectively. In the initial training phase, we started with fully connected CNNs and U-nets (Ronneberger et al., 2015) and by incrementally adding datasets to the training set. The first goal of the model was to boost salt edge detection with high accuracy. High accuracy has two components:

1. Detect salt with high overlap with the labels, and
2. The position of the detected edge consistently aligns with the seismic loops.

For top-salt and TSOH models, the salt edge should consistently align with the positive peaks in the seismic signal, whereas for base-salt and BSOH models the salt edge should consistently align with the negative peaks in the seismic signal.

Given these established criteria for the DL models, we incrementally started to incorporate more data from more diverse surveys. Training the model with more data helped in boosting the salt edge detection along areas with strong seismic signature; furthermore, adding more diverse data also improved the generalization ability of the model. However, we also noticed two degradations in the quality of detections:

1. Consistent increase in the persistent false salt detection (or false positives) and they increased with the amount of data added to training. An example of the false detections is highlighted in the yellow ovals in Figure 4, panel b.
2. Incorrect positioning of some of the salt detection, especially in areas of complex geology and areas of low SNR in the seismic, like overhangs. An example of the incorrectly positioned detections is highlighted in the black circle in Figure 4, panel b, as compared to the salt label in panel a.
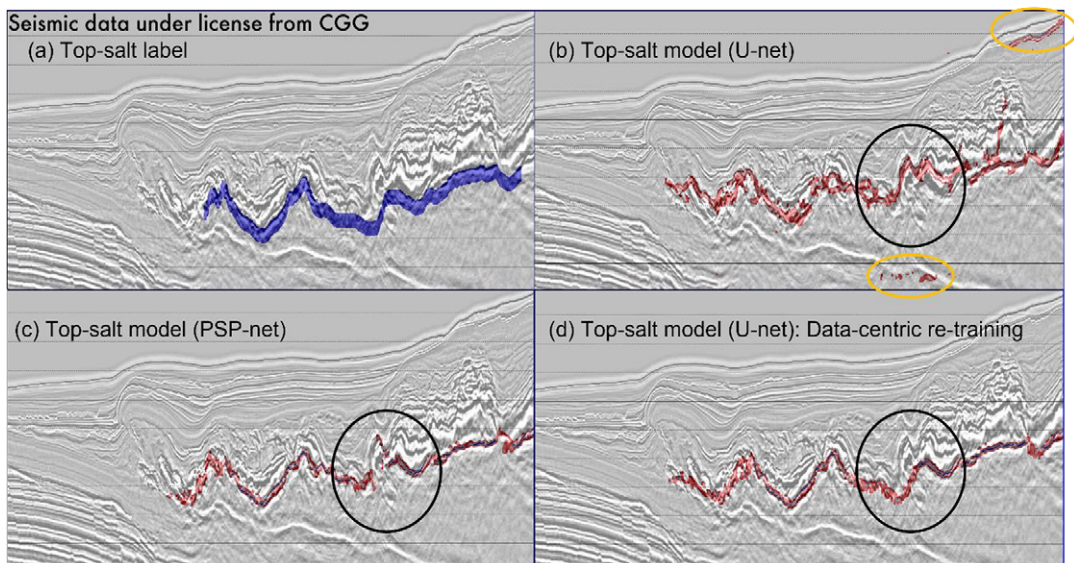

Figure 4. *The figure shows a comparison of multiple top-salt models (c and d) re-trained using model-centric and data-centric methodologies to improve top-salt (b) detection inaccuracies.*
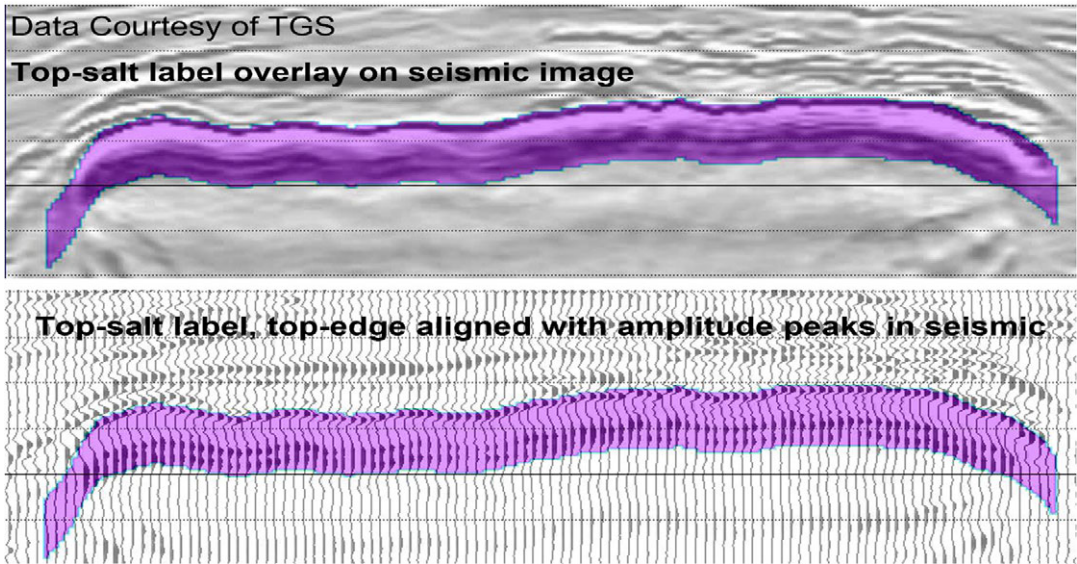
Figure 5. *The top image shows the overlay of the thick top-edge label used for training, and the bottom image shows the same label to highlight the top-edge of label is aligned with seismic peaks.*

To address the false salt detection, the number of 2D training tiles where no salt exists, hence the no label, were added to the training data used to train the models. In general, the number of non-salt used was about a forth the total number of tiles with salt. This provided a reasonable balance between 2D tiles of positive and negative examples of salt detection for the model. This strategy reduced the deeper false salt detections significantly, yet the shallower ones persisted.

To address the complexity in the geology, we introduced more sophisticated neural network architectures like, fusion net (Quan et al., 2016), dense net (Huang et al., 2017), and pyramid scene parsing (PSP) networks (Zhao et al., 2017). Simply going from U-net to PSP architecture does not allow to train a model that addresses the in-correct salt detection, as is evident from panels b and c of Figure 5, highlighted by the black circle. However, going to a more complex architecture did remove the shallow false detections, at the same time resulted in breaks in the detected edge as well.

As mentioned in the introduction section, another requirement of the DL models in terms of model performance is that they should be integrated into the existing salt interpretation workflow. This requirement renders model's performance characterized by traditional metrics like accuracy is less meaningful. Rather the model's performance at creating an inference best suited to maximize the performance of any downstream tasks in the workflow is most crucial. In the context, the inference generated by the trained DL model needs to be converted into a 3D surface data structure. Hence, for the process of conversion of inference to 3D surface to be effective, it is more important that the DL model produce as little false detections while being highly accurate in coverage and positioning of the detected edge.

Even with increasing the amount and diversity of training data and adopting more complex neural network architectures, the false salt detections in incorrect positioning of detected salt persisted making the models well-generalized yet unsuitable for production.

4. Data-centric strategies to address DL model challenges

From the observations from model-centric experiments, we systematically explored the relationship between the dataset, data quality (seismic quality and label quality), and the performance of the DL model

trained using that data. We concluded that the following characteristics of the training data are essential for effective training.

4.1. Label quality

The salt interpretations that are used as labels are typically single-valued edges along seismic peaks. We found that these labels were more susceptible to errors in the labeling process. The labels were hence made thicker, as shown in Figure 5, to alleviate incorrect or imprecise labels. We experimented with several label thickness values and determined that any thickness value that reliably encompasses one to two adjacent seismic loops boosts robustness of prediction positioning and better localization on the loop of interest. This also acts as implicit regularization and helps generate wide range of prediction probabilities, mainly in areas of ambiguity in the seismic edges.

4.2. Seismic quality along salt edge

We observed that there were two areas where the DL model struggles, either to detect salt resulting in loss of salt coverage or incorrectly positions salt.

1. Areas of complex geology, which manifests as ambiguous seismic signature, as highlighted by black circle in panel b in Figure 4.
2. Areas of weak or no seismic signature, as highlighted by the red circles in panels c and d in Figure 6.

In the model-centric approach described in Section 3, for a given 3D seismic volume, all the tiles created from that volume where salt label exists, in other words, salt exists, are used for training. In response to the above observations, we adopted a slightly different strategy during the selection of training tiles, tiles where geology was sufficiently complex or seismic signature was weak/non-existent but contained salt labels were deliberately excluded from the training.

As the percentage of data with weak or ambiguous seismic signature for salt labels reduces, the position accuracy of trained model in complex geology or ambiguous seismic improved. This improvement in positioning is illustrated within black circle in panel d of Figure 4. Within the black circle, neither the
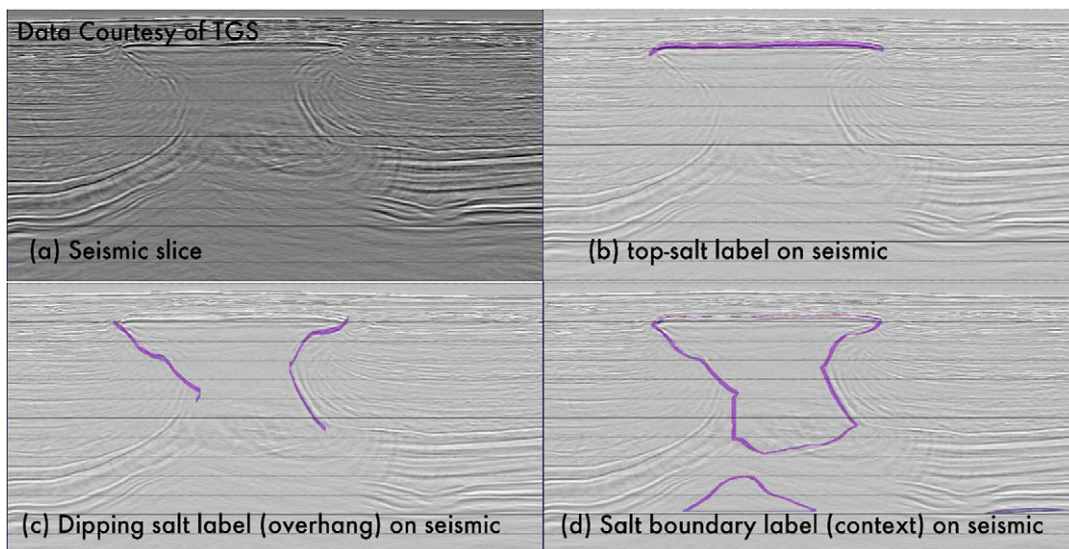

Figure 6. *Figure shows the seismic panel with different salt edge labels overlaid, images c and d showcase the salt overhang labels and the complete salt boundary labels to highlight the structural context which enables DL models to learn effectively.*

U-net (panel b) or PSP-net (panel c) produces a detection aligned with the label (panel b). With data selection based on seismic quality and increasing thickness of labels, even with comparatively less complex architecture like U-net the salt edge positioning is accurately placed.

In addition, in other complex areas, the model simply did not generate detections instead of generating erroneous ones. Thus, removing data where the labels were not supported with clear seismic signatures from training data enables the model to learn the salt-sediment boundary effectively, as the model is more robust to ambiguous seismic as it relates to interpretations (or labels).

To operationalize this strategy in the training data generation pipeline, we filter out tiles where the labels are significantly inconsistent with the seismic loops or entropy of the seismic data underneath the label is below an acceptable threshold. This also generated models that are robust to small perturbation in seismic alignment and the presence of noise, thus data with smaller alignment deviations could still be used to reliably train. This allowed us to leverage quite significant sections of large seismic surveys into training. In addition, we pruned out the tiles that only have a smaller salt percentage within the tile.

4.3. Sufficient structural context

However, the detection of salt where seismic signature is weak or non-existent persisted. For models trained to detect steeply dipping edges of the salt body (called overhangs) which are not very common and tend to be poorly imaged, very few representative datasets with strong seismic signature were available.

Even though the salt overhang labels were not entirely supported with well-imaged seismic, if we leverage labels that outline the entire salt boundary, which we refer to as salt rings, within a tile, there will be enough area of the label that has a strong seismic support. Such labels also provide a structural context to the salt boundary. For the overhang specific DL models, the labels were extended to include not just the steeply dipping edges but the entire salt boundary. Panels c and d from Figure 6 show a visual comparison of steeply dipping salt edge labels with structurally more complete salt boundary labels.

Just a few datasets, with enhanced salt structure context labels along with standard data augmentation techniques, was a significant breakthrough in effective model training. Figure 7 Illustrates the
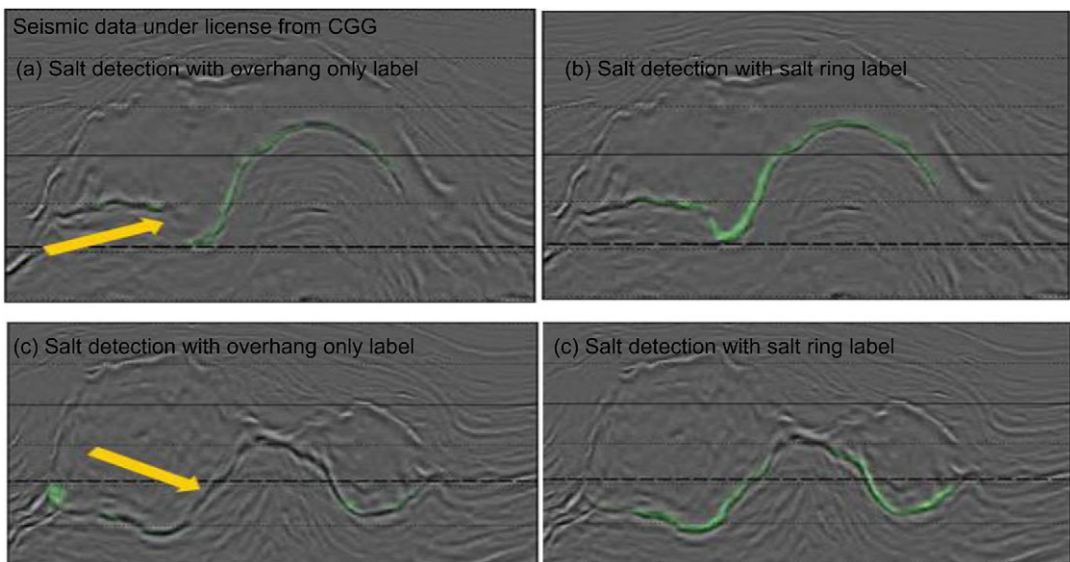

Figure 7. *Figure shows two pairs of seismic panels with salt overhang detection coverage (panels a and b) of DL model trained using only overhang labels compared to model (panels c and d) trained using salt ring labels. It showcases the salt overhang coverage is boosted significantly using salt ring labels to train the models.*

enhancement in the DL models' performance after training with entire salt ring labels. The yellow arrows in panels a and c point to the areas of the overhang that the original DL model was unable to detect the salt edge; however, if the DL model is trained with sat ring, which embeds the salt's structural context, the overhang coverage is boosted significantly. Better coverage on the overhang makes the downstream task of 3D surface extraction also simple to complete with high accuracy.

Thus, enhancing the quality of the training data by systematically focusing on training data generation that boosts label context, reduces label ambiguity, and increases seismic support for the label improved the model performance. The focus on optimization of the training data and label quality to improve the DL model's performance makes these strategies data-centric rather than model centric and enables the well-generalized model to be deployed in production and integrated into the salt interpretation workflow.

The DL models, which turned out to be the most robust and suitable for deployment, were trained with much less training data still encompassing the geological variability, however, with much higher quality. In this context, high quality refers to the data that allow the models to learn effectively; in other words, data that allow the model to learn robust representations of the task it is learning. For salt interpretation, these observations were incorporated into automated training data generation through systematic label generation and data quality filters.

5. Data-centric strategies for sustained DL model deployment

As DL models are deployed, it is essential for sustained deployment that:

1. DL models are interpretable:
    The users must understand the model's behavior, prediction performance, and failure modes. This understanding helps the user better explain model outcomes, builds user trust, and hence boosts optimal DL model usage.
2. Models are continuously re-trained:
    As DL models are deployed, they encounter data drift. Data drift is the change in the distribution of data the DL model is exposed to overtime. Consequently, model's inference quality begins to degrade, as test data begin to deviate from the training data, which is inevitable in real-world settings. With seismic data, it is quite challenging to quantify the distribution of the training dataset and even more so of the test dataset. Thus, quantification and subsequent combating of data drift extremely difficult, directly impacting the longevity of the deployed models. To maintain acceptable performance of the DL models, continuous improvement of models through re-training by adding production to the training set needs to be enabled.

5.1. Data quality criterion for model performance prediction

Most of the salt interpretation DL models were trained on legacy datasets and during deployment, applied to data from new terrains which are acquired and processed through the most recent acquisition and processing techniques. This further adds to widening data drift the salt interpretation models encounter.

Given the data-centric strategies applied to generate the training data, it is reasonable to expect that data where the quality of seismic image at salt boundary is strong, the DL models will perform well, in other words, high coverage and accurate positioning of detected salt. In addition, salt interpretation was posed as an edge detection problem, so the DL models were optimized to detect the salt edges. Hence, it follows that the model will be better able to delineate the salt edge where the edge contrast is high. Based on this reasoning, we heuristically designed a set of criteria that quantify the minimum data quality required to perform the model's task manually.

This criteria, called data conditioning criteria, consists of two parts: one is the lower bound on seismic data contrast (quantified using RMS contrast) and second, detects seismic loop clipping. The data conditioning criteria has served in practice, as a reliable predictor of a trained model's performance on a blind dataset, as it roughly quantifies whether the test data sufficiently capture phenomena that model

needs to generate acceptable performance. In addition, the criteria also provides the interpreter, with guidance to improve the blind/test data quality through processing techniques that will allow the salt interpretation DL model to better delineate the salt-sediment boundary, in effect, providing a data quality sufficiency test.

The ability to predict the DL model's performance and guidance to improve data quality both contribute significantly to boost the interpretability of the DL model.

Figure 8 shows the evolution in the seismic contrast and seismic loop clipping, as the seismic is processed with different filters. As is evident as the seismic processing diminishes the seismic signature along salt edges, we see the RMS values evolve and detected salt coverage is also affected adversely.

Furthermore, the data conditioning criteria allow to loosely detect a domain gap as the model is deployed in real world. Domain gap refers to the large difference between the model training data and the model test data and is quite difficult to understand. Thus, the metric can potentially be leveraged to automate the selection of high-quality data for continuous improvement of models in production. In effect, providing a lower bound for training appropriateness analysis of new datasets and enabling re-training of DL models to boost longevity.

5.2. *Statistical data quality testing to predict model failure modes*

The data-centric training of DL models has underscored the importance of seismic data quality in developing reliable, well-generalized models. Hence, it follows that DL model's salt detection performance is also highly correlated with seismic quality, particularly as it related to salt edges or salt-sediment boundaries. As seismic quality diminishes, the model performance will also degrade, and to better understand the limits of the DL model performance, we implemented statistical tests, which aim to quantify an "operating range" of seismic quality beyond which the model's seized to detect salt. This in effect provides the failure cases for the DL models.

The statistical tests are designed, given an indication of how robust the DL model performance is to statistical variations in test data. For these statistical tests, a set of blind datasets that vary from the training data in terms of acquisition and processing techniques were selected as the test corpus. Subsequently, these datasets were subjected to a series of common seismic processing algorithms to simulate effects real-world processing on quality of seismic imaging. Figure 8 shows the 2D slices of the original seismic data and the corresponding processed slices with two different processing techniques. This provided a set of data that represent the variations in seismic quality that the models may encounter in real-world settings. The salt interpretation model's performance over these seismic variations was used to empirically
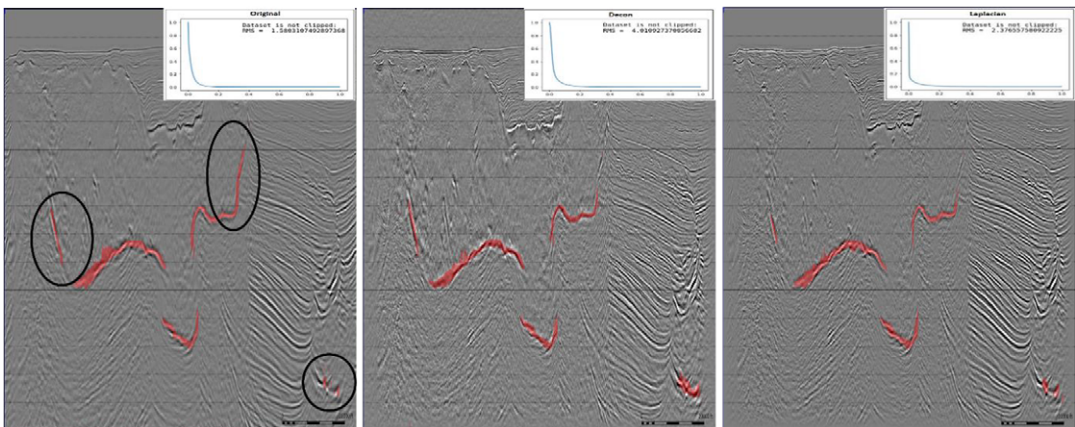

Figure 8. *The panels from left to right showcase the original seismic data, de-convolved seismic and Laplacian filtered seismic. The plots on top right of each panel shows the evolution of the data-conditioning criterion associated with each seismic.*

determine a set of processing algorithms and the associated parameters which diminished the model's performance beyond the acceptable range. These tests further boost the DL model's interpretability by determining an "operating range" where the model will provide reasonably accurate salt detection.

6. Lessons learned

The inherent geological variability and ever-evolving quality of seismic data coupled with the complexity and nature of salt were the major challenges in developing salt interpretation DL models. The popular model centric strategies for traditional computer vision models fell short in overcoming this challenge. It was only when we started to account for the fundamental and domain-specific distinctions in type and nature of seismic data compared to natural images in the training data generation pipelines that the salt interpretation model performance started to improve, making the models well-generalized and suitable for production. In other words, data-centric strategies proved extremely effective in incorporating domain understanding into DL model development.

Strong seismic signatures and proper alignment of label with these signatures proved essential for effective training of salt interpretation models. Only the labels which were supported by good quality seismic consistently boosted the model's ability to improve salt detection coverage with high positional accuracy at the same time maintain good generalizability. The models were highly sensitive to incorporation of data where the seismic support of labels was either ambiguous (complex geology) or poor quality (poor image or too deep).

Adding training data where no salt is available, or negative examples, was effective to consistently address false positives, and strategy was key in ensuring that the DL model's inference has minimal false detections for the downstream tasks in the salt interpretation process to be effective.

These data-centric insights can also be favorably leveraged either through data quality quantification or statistical manipulation of data quality to boost DL model's interpretability and continuous improvement to enhance model's longevity in deployment, consequently realizing long-term value.

In conclusion, this paper summarizes our empirical observations which suggest that training DL models for real-world, domain-specific problems that are generalizable and robust to noise and/or data artifacts requires careful selection of training data and labels. Based on computer vision literature, the data-centric strategies are quite intuitive, in that for the model to learn the relationship between the data descriptors and task at hand, good quality data descriptors are essential. However, principled insights to answer two questions: How do we characterize "quality" that allows DL model to train effectively? and How do we quantify this quality? are quite difficult to conclude, and this is the main data engineering challenge for automation of data-centric AI strategies.

References

Alkan E, **Cai Y**, **Devarakota P**, **Gala A**, **Kimbro J**, **Knott D**, **Madiba G, and Moore J** (2022). *SaltCrawler: AI solution for accelerating velocity model building*. SEG Technical Program Expanded Abstracts: 1684–1688.

Cubuk ED, *et al.* (2020). *Randaugment: Practical automated data augmentation with a reduced search space.* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

DeVries T (2017). Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552.

Gala A, *et al.* (2022). *Towards continual AI for salt interpretation using Incremental learning.* SEG Workshop on Applications of ML and AI in Geophysics.

Huang G, *et al.* (2017). *Densely connected convolutional networks.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Quan TM, *et al.* (2016)*. FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Ronneberger O, *et al.* (2015). *U-Net: Convolutional networks for biomedical image segmentation*. Medical Image Computing and Computer Assisted Intervention-MICCAI.

Shi Y, **Wu X and Fomel S** (2019). SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network. Special section: Machine learning in seismic data analysis, *Interpretation*, *7*, SE113–SE122.

Zhao H, *et al.* (2017). *Pyramid Scene Parsing Network.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.