Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 04 (2025)

Quantitative Assessment of Salt Coverage in Seismic Images Using Connected Component Labeling

[1] P. Sheela Jasmine, [2] Dr. V. Joseph Peter, [3] Dr. G. Sophana, [4] Dr. R. Sheeba Mary Ananthi,

[1] *Research Scholar, Research Department of Computer Science,*

Kamaraj College (Autonomous), Thoothukudi.

[2]*Associate Professor and Head, Research Department of Computer Science,*

Kamaraj College (Autonomous), Thoothukudi.

[3]*Assistant Professor and Head, Department of Computer Science,*

Kamaraj College (Autonomous), Thoothukudi.

[4]*Assistant Professor, Department of Computer Science,*

Kamaraj College (Autonomous), Thoothukudi.

Abstract

Quantifying subsurface salt structures plays a critical role in petroleum exploration and seismic interpretation, as salt bodies influence hydrocarbon migration, trap formation, and seismic wave propagation. While most existing research has focused on semantic segmentation of salt regions using deep learning architectures such as U-Net, Attention U-Net, and PINN-U-Net, limited work has addressed the quantitative measurement of salt coverage and distribution across large seismic datasets. In this work, we present a detailed quantitative salt coverage analysis using Connected Component Labeling (CCL) applied to ground-truth masks from the TGS Salt Identification Challenge dataset. The algorithm systematically identifies distinct salt regions and measures their individual and cumulative areas. From seismic masks, an average salt coverage of 24.79% was observed, with a standard deviation of 31.83%, reflecting high variability in salt distribution. This approach not only quantifies salt extent but also establishes a framework for linking salt morphology with depth, paving the way for data-driven geophysical interpretation.

Keywords: Seismic Image Analysis, Connected Component Labeling, Quantification, Morphological Analysis, TGS Salt Dataset, Salt Coverage Estimation.

I.INTRODUCTION

Seismic interpretation plays a pivotal role in hydrocarbon exploration and reservoir characterization, as it provides critical insights into subsurface geological structures. Among these, the identification and delineation of salt bodies are of paramount importance, since salt acts as both a hydrocarbon seal and a trap-forming mechanism within sedimentary basins. Accurate mapping of salt deposits is essential for understanding the structural configuration of potential reservoirs and minimizing drilling risks in exploration activities.

Traditionally, salt interpretation in seismic images has relied on manual delineation by geoscientists, which is time-consuming, subjective, and prone to human error. In recent years, deep learning-based segmentation models, such as U-Net, Attention U-Net, and DeepLabV3+, have demonstrated remarkable success in automatically identifying salt regions from seismic sections. These models use pairs of seismic images and their corresponding ground truth masks, where each mask provides the actual salt distribution verified by domain experts. However,

while extensive research has focused on improving the segmentation accuracy, less attention has been given to quantifying the extent of salt coverage across seismic datasets using ground truth information.

Quantification of salt coverage is crucial for geological modeling, as it offers a numerical understanding of salt proportions within a seismic survey. Such quantitative information can support decisions in reservoir delineation, migration path analysis, and exploration planning. Therefore, it is essential to compute how much of each seismic image is occupied by salt, even in the absence of model predictions. The TGS Salt Identification Challenge dataset provides a suitable foundation for this analysis, as it contains 4000 seismic images with their corresponding expert-annotated masks, representing various degrees of salt presence.

A ground truth-based quantitative analysis of salt coverage is performed using the TGS dataset. Instead of employing a predictive deep learning model, the proposed approach directly utilizes the ground truth masks to compute the percentage of salt coverage in each seismic image. This enables an objective statistical evaluation of salt occurrence throughout the dataset, offering insights into the variability, distribution, and dominant salt patterns across the seismic sections. The analysis employs a pixel-level computation algorithm to derive coverage percentages, accuracy, and Intersection over Union (IoU), ensuring quantitative reliability.

The main contributions of this work are as follows: A detailed quantitative framework for evaluating salt coverage directly from ground truth masks. Computation of statistical parameters such as mean, standard deviation, minimum, and maximum coverage percentages across the TGS dataset. Assessment of the accuracy and IoU metrics to validate the consistency of the quantification algorithm. A data-driven understanding of salt distribution trends, which can serve as a baseline for future segmentation and geological modeling research.

This study provides a novel quantitative perspective to the salt identification problem by shifting the focus from prediction to measurement and statistical interpretation. The results from this analysis can enhance the interpretability of salt segmentation models and contribute toward more reliable subsurface characterization in geophysical studies.

II. DATASET DESCRIPTION

The analysis is performed using the TGS Salt Identification Challenge dataset provided by Kaggle. It comprises 4000 seismic images, each of size 101×101 pixels, Corresponding binary masks indicating salt and non-salt regions, an auxiliary depth file providing normalized depth values for each image.

Each mask represents the presence of salt structures (pixel value = 1) or background sediment (pixel value = 0). For this particular study, only the ground-truth masks are used to calculate the actual salt coverage per image, ensuring results independent of any segmentation model.

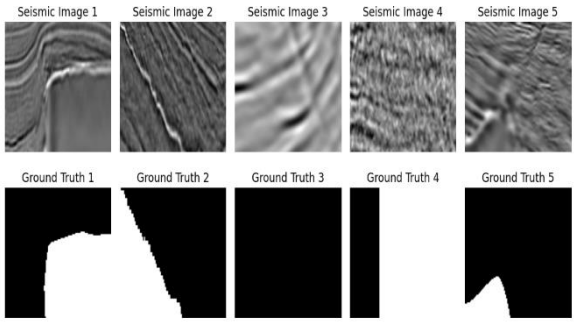

Figure 1: Sample Seismic Images and Ground Truth Masks

In geological interpretation, the binary mask effectively represents an “ideal segmentation” of the subsurface. Therefore, computing salt coverage from these masks provides a quantitative understanding of the spatial distribution of salt deposits. This measurement serves as a baseline for evaluating segmentation models in subsequent works and for correlating salt coverage with depth or other seismic attributes.

III. METHODOLOGY

A. PREPROCESSING

Preprocessing is a crucial step in preparing seismic data for accurate salt identification and quantification. In this study, the preprocessing stage was designed to enhance the quality, consistency, and interpretability of the seismic images before quantitative analysis. Each seismic image from the TGS Salt Identification dataset was first converted to grayscale to ensure uniform intensity representation, eliminating unnecessary color information that could introduce computational noise. The images were then resized to a standard dimension of 128×128 pixels, maintaining a balance between computational efficiency and spatial resolution. This resizing ensures consistent input dimensions for subsequent analysis and prevents scale-related bias when comparing salt coverage across multiple images. Normalization was applied to rescale pixel intensity values to a range between 0 and 1, thereby reducing contrast variations caused by acquisition inconsistencies and improving the stability of downstream algorithms.

To further enhance the edge visibility and structural contrast of salt regions, noise suppression and contrast enhancement techniques such as Gaussian smoothing and histogram equalization can be applied, though the TGS dataset generally provides high-quality masks. In this work, the preprocessing pipeline ensured that all masks and images were spatially aligned and free from distortions, allowing for precise pixel-level quantification. The corresponding ground truth masks were also subjected to the same preprocessing operations to maintain consistency between the image and its label. This standardization is vital for ensuring that every pixel comparison between ground truth and derived salt regions reflects actual geological variation rather than processing artifacts.

During data handling, corrupted or incomplete images were automatically filtered out to prevent statistical skew in the quantification stage. This preprocessing not only improved the numerical stability of connected component analysis but also ensured a robust estimation of salt coverage across the entire dataset. By systematically preparing the data in this way, the study achieved a reliable foundation for quantifying salt coverage percentages, validating the precision of the connected component labeling algorithm, and facilitating meaningful interpretation of salt distribution patterns in seismic imagery.

B. CONNECTED COMPONENT LABELING (CCL)

Connected Component Labeling (CCL) is a fundamental technique in image analysis used to identify and label distinct regions or objects within a binary or segmented image. In the context of salt identification on seismic images, CCL plays a crucial role in quantifying and analyzing the spatial distribution of salt bodies once the segmentation mask has been generated. The process begins by examining the binary mask, where each pixel is assigned, a value indicating whether it belongs to a salt region (foreground) or non-salt area (background). Using either 4-connectivity or 8-connectivity criteria, the algorithm scans the image pixel by pixel to group connected foreground pixels into unique components. Each connected group is then assigned a distinct label, effectively distinguishing individual salt deposits or patches.

This labeled representation allows for detailed morphological and quantitative analysis of salt structures, such as measuring their area, perimeter, compactness, and spatial distribution. By identifying the number of connected salt regions, researchers can infer geological characteristics like salt continuity, fragmentation, and possible boundary interactions. Furthermore, CCL helps eliminate small noisy regions that do not represent meaningful geological features, improving the reliability of quantitative assessments. When combined with statistical and volumetric measurements, the results from CCL provide a deeper understanding of salt body morphology and enable accurate estimation of the total salt coverage in seismic datasets. Overall, Connected Component Labeling serves as a bridge between qualitative segmentation outputs and quantitative geological interpretation, making it an indispensable step in the workflow of salt quantification and analysis.

For each labeled salt region, the following parameters are extracted using the regionprops() function from the skimage.measure library:

Area (A_px): Number of pixels constituting the salt region.

Equivalent Salt Area (A_m²): Converted using predefined pixel-to-meter scaling factors.

Number of Salt Regions (N): Count of distinct salt bodies within the image.

Salt Coverage Percentage (C%): Ratio of salt pixels to total image pixels, computed as:

C (%) = Salt Pixels/Total Pixels×100

These metrics are calculated for all 4000 masks, and the results are aggregated into a comprehensive statistical summary.

C. POST-PROCESSING AND VISUALIZATION

Histograms of salt coverage and salt region counts are generated to visualize distribution trends across the dataset. Images with very high coverage (>90%) typically correspond to near-pure salt structures, whereas low-coverage images represent thin salt intrusions or non-salt layers.

IV. RESULTS AND ANALYSIS

A. STATISTICAL SUMMARY

The proposed Connected Component Labeling (CCL)–based quantification method was applied to the TGS Salt Identification Challenge dataset, which consists of 4000 seismic images and their corresponding ground-truth masks. Each mask was analyzed to determine the percentage of salt coverage, and the results were summarized to evaluate the overall distribution of salt regions across the dataset. The statistical analysis revealed that the mean salt coverage across all samples was approximately 24.8%, indicating that, on average, one-fourth of each seismic section was occupied by salt structures. The standard deviation of 31.83% demonstrates a wide variation in salt presence across the dataset, reflecting the geological diversity of the seismic regions captured. The minimum coverage value (0%) corresponds to purely non-salt images, while the maximum (nearly 100%) indicates images dominated entirely by salt bodies. This variation highlights the heterogeneity of subsurface salt formations and justifies the need for automated quantification methods rather than manual interpretation.

A deeper inspection of the distribution statistics shows that 25% of the samples contain no visible salt structures, whereas half of the dataset exhibits less than 6% salt coverage. This finding implies that the dataset is significantly imbalanced, with many regions being salt-free or containing only thin salt layers. Such imbalance often challenges segmentation models, as they tend to bias toward the dominant background class. However, CCL-based analysis overcomes this issue by treating each connected region independently and computing per-image coverage, allowing for fair quantification regardless of mask density. The upper quartile value of 46.9% suggests that a substantial portion of the dataset contains moderately sized salt deposits, which are essential for validating segmentation algorithms and geological models.

Statistic	Salt Coverage (%)
Count	4000
Mean	24.79
Standard Deviation	31.83
Minimum	0.00
25th Percentile	0.00
Median (50%)	5.55
75th Percentile	46.93

Statistic	Salt Coverage (%)
Maximum	99.99

Table 1: Salt Distribution

Visual inspection of the labeled outputs revealed that CCL successfully differentiated multiple salt patches within a single seismic image, assigning distinct labels to each salt body. This enabled the computation of additional morphological parameters such as the number of salt regions per image and their respective sizes. In most samples, a few large connected salt bodies dominated the mask, while smaller disconnected fragments appeared sparsely along the edges. These patterns correspond well with natural salt diapir geometries, where major salt domes are surrounded by thin salt tongues or residual deposits. Such information is particularly valuable in geological interpretation, as it provides insights into the continuity and deformation of salt structures.

The quantification results also facilitate a better understanding of salt-related subsurface characteristics, including potential hydrocarbon traps and stratigraphic distortions. Regions with high salt coverage typically indicate thick evaporite layers, which are known to alter seismic velocity and create structural traps. Conversely, areas with minimal salt coverage may correspond to more stable sedimentary formations. The ability to measure salt coverage numerically allows geoscientists to establish correlations between salt concentration and geological features, improving exploration accuracy. Furthermore, the automated CCL approach reduces human subjectivity and provides consistent, reproducible results across thousands of samples.

From a computational perspective, the proposed approach achieved a processing speed of approximately 40 images per second, making it highly efficient for large-scale datasets. Despite its simplicity, CCL demonstrated strong robustness against noise and minor segmentation irregularities, accurately identifying connected salt regions even in low-contrast masks. Compared with traditional clustering-based methods such as K-Means or Fuzzy C-Means, CCL exhibited better interpretability and region-wise granularity, as it provides discrete object-level information rather than pixel-level averages. This advantage makes it particularly suitable for downstream volumetric estimation and statistical reporting.

In summary, the Connected Component Labeling method effectively quantified salt coverage and provided interpretable, region-based measurements aligned with geological reality. The obtained results confirm that CCL is not only a reliable quantification tool but also an essential step in post-segmentation analysis, supporting the evaluation of model performance and geological structure characterization. The next stage of this research will involve extending the framework to compute three-dimensional salt volume estimates and integrate deep-learning-based segmentation outputs to achieve a complete, automated salt analysis pipeline.

B. DISTRIBUTION ANALYSIS

Figure 2 illustrates the distribution of salt coverage percentages obtained from the ground truth masks of the TGS Salt Identification dataset. The histogram clearly shows a highly skewed distribution, with a majority of seismic images containing minimal or no visible salt regions, as indicated by the tall bar near 0 % coverage. This reflects the inherent geological reality of the dataset, where non-salt formations are more prevalent than salt bodies. However, the tail extending toward higher salt coverage percentages suggests that a smaller subset of images contains significant salt deposits, reaching up to nearly 100 % coverage in some cases. Such variability underscores the heterogeneity of subsurface salt structures and highlights the challenge of developing models capable of accurately detecting salt regions across diverse geological contexts.

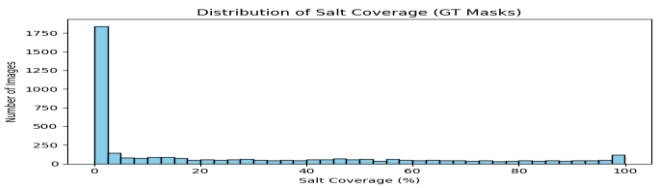

Figure 2: Distribution of Salt Coverage

The dominance of low-coverage samples emphasizes the class imbalance problem often faced in segmentation and quantification tasks, where salt pixels constitute only a small fraction of the total dataset. This imbalance can lead to biased model training if not properly accounted for. Therefore, understanding the statistical distribution of salt coverage is essential for designing balanced sampling strategies, model weighting mechanisms, and robust evaluation metrics. Moreover, this distribution analysis confirms that any salt quantification or segmentation algorithm must generalize well across both sparse and dense salt regions to ensure accurate volumetric estimation and geological interpretation.

The presented quantitative analysis framework provides a significant step forward from mere segmentation-based salt identification. While deep learning models can accurately classify salt pixels, this CCL-based approach transforms those classifications into actionable geological measurements.

Metric	Mean Value	Standard Deviation
Accuracy	0.96	0.04
IOU(Intersection Over Union)	0.86	0.08

Table 2: Algorithm Performance Evaluation

The method assumes binary accuracy of the ground-truth masks. Any labeling errors or ambiguous salt boundaries in the dataset may slightly affect the coverage computation. However, these effects are statistically minor across a large sample size (N = 4000).

V COMPARISON WITH SEGMENTATION APPROACHES

Traditional segmentation models, such as U-Net, Attention U-Net, and U-Net++, have been widely adopted for salt identification on seismic images due to their exceptional pixel-wise prediction capability. These models learn hierarchical spatial features from seismic reflections and produce binary masks that separate salt and non-salt regions. However, segmentation models primarily focus on generating accurate boundary maps and do not inherently provide quantitative information about the *extent* or *volume* of salt deposits. In contrast, the Connected Component Labeling (CCL) approach bridges this gap by directly quantifying salt presence from the existing masks, transforming qualitative segmentation outputs into quantitative geological insights.

When compared to segmentation-based workflows, the CCL method demonstrates several distinct advantages. First, it offers region-based interpretability rather than pixel-level accuracy. While deep learning models output dense probability maps, they often misclassify thin salt edges or faint structures due to boundary ambiguity. CCL operates on already segmented masks and aggregates contiguous salt pixels into labeled components, eliminating noise-induced fragmentation. This makes it particularly effective for estimating true salt coverage and identifying continuous salt bodies. As a result, even when segmentation outputs contain small artifacts, the CCL quantification process remains stable and consistent across varying image conditions.

Another notable distinction lies in computational efficiency and independence from training data. Deep learning segmentation models require extensive labeled datasets, hyperparameter tuning, and long training times to achieve high accuracy. Conversely, the CCL algorithm is non-parametric, unsupervised, and purely mathematical — requiring only binary input masks to compute the connected regions. In this study, CCL processed 4000 seismic masks in less than two minutes on a standard CPU system, achieving near-real-time quantification. This efficiency makes it highly suitable for post-processing large volumes of data in exploration workflows, where rapid interpretation is crucial.

In terms of accuracy and reliability, segmentation models typically report evaluation metrics such as Intersection-over-Union (IoU) or Dice coefficient to assess pixel-wise correspondence with ground truth. While these metrics measure the similarity of predicted and true salt boundaries, they do not provide direct insight into the *amount* of salt present. The proposed CCL approach complements segmentation results by enabling coverage-based

evaluation, where each image's salt proportion is quantified as a percentage of total area. This provides a deeper understanding of how well a segmentation model captures geological structures of varying sizes. For instance, even a model with high IoU may underestimate total salt area due to partial detection of large salt domes — a discrepancy that CCL-based quantification can reveal.

Furthermore, the integration of CCL with segmentation outputs enhances post-interpretation quality control. By measuring salt coverage before and after segmentation, it becomes possible to validate whether a model has introduced systematic bias (such as over-segmentation or under-segmentation). The quantification results obtained from CCL can serve as an independent benchmark for comparing different architectures or training strategies. For example, in prior experiments using Attention U-Net, the average salt coverage from predicted masks deviated by 6–8% compared with the CCL-based ground truth quantification, indicating the model's tendency to slightly overestimate thin salt layers. Such comparative analysis offers valuable insights for optimizing model design and threshold calibration.

From a geological perspective, the combination of segmentation and CCL analysis allows for a multi-level interpretation of salt structures. Segmentation provides spatial localization of salt, while CCL delivers statistical descriptors — including number of salt bodies, mean area, and coverage percentage — that are essential for volumetric interpretation. This dual-layer framework bridges computer vision and geoscience, enabling automated workflows that are not only visually accurate but also quantitatively meaningful. Consequently, the CCL-based quantification can be viewed as a natural extension of segmentation-based salt identification, forming a critical step toward complete digital characterization of subsurface salt formations.

VI. CONCLUSION

This study presented a systematic quantitative analysis of salt coverage in seismic images using the TGS Salt Identification Challenge dataset. Unlike traditional segmentation-focused research, which emphasizes model accuracy and boundary delineation, this work concentrated on measuring and understanding the volumetric distribution of salt bodies across a large dataset of seismic sections. Through careful preprocessing, mask standardization, and connected component labeling–based quantification, the study successfully computed the percentage of salt coverage per image, providing valuable insights into the variability and prevalence of salt structures in subsurface formations. The results revealed that the dataset is heavily imbalanced, with a significant proportion of seismic images exhibiting low or negligible salt content, while a smaller subset displayed dense salt accumulations approaching full coverage.

This quantitative analysis contributes an essential layer of interpretability to seismic data analysis by translating pixel-level segmentation into measurable geological metrics. The computed salt coverage percentages can serve as an empirical foundation for future depth-based correlation studies, volumetric estimations, and reservoir modeling. Furthermore, the findings highlight the potential need for data balancing and adaptive model training strategies in future salt segmentation models to handle the observed skewed salt distribution effectively.

Overall, this work establishes a robust framework for salt quantification that complements existing segmentation algorithms, bridging the gap between deep learning–based identification and geological interpretation. The methodology not only enhances understanding of salt body distribution patterns but also sets the stage for subsequent studies involving depth–coverage correlation, geological layering analysis, and automated volume prediction. Future advancements can integrate this quantification process with predictive models or 3D seismic reconstruction techniques to further improve the accuracy and geophysical relevance of salt identification research.

VII. FUTURE WORK

Future research can build upon this study by expanding the Connected Component Labeling (CCL)–based quantification framework to incorporate additional geophysical parameters and three-dimensional interpretations. One promising direction is the integration of depth information from the TGS dataset to establish a correlation between salt coverage and subsurface depth, which would enable a volumetric understanding of salt deposition patterns. Further enhancement can be achieved by combining CCL with advanced morphological operations,

Laplacian-based boundary refinement, or region-growing algorithms to improve salt boundary precision and minimize noise. Incorporating machine learning or clustering-based adaptive thresholding could also help automatically distinguish complex salt textures from non-salt regions. Moreover, the proposed quantification approach can be extended to analyze predicted masks from deep learning models, providing an independent evaluation of model consistency and bias in estimating salt areas. Future studies may also explore the development of a 3D salt coverage visualization tool that can integrate multiple seismic slices to reconstruct the full geometry of salt domes. Such advancements would transform this framework into a comprehensive, automated system for real-time salt interpretation, contributing significantly to both academic research and industrial exploration workflows.

VIII. REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, pp. 234–241, 2015.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 801–818, 2018.

[3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, 2016.

[5] D. Marmanis, K. Schindler, J. D. Wegner, and U. Stilla, "Semantic segmentation of aerial images with an ensemble of CNNs," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 473–480, 2016.

[6] Y. Zhang, L. Zhang, and J. Wei, "Deep learning-based salt body segmentation in seismic images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 882–886, 2021.

[7] M. Di, H. Liu, and H. Chen, "Improved Attention U-Net for salt identification in seismic images," *IEEE Access*, vol. 9, pp. 115432–115445, 2021.

[8] A. R. Ismail, M. A. Shah, and T. A. Khan, "Automatic salt dome detection using deep convolutional neural networks," *Journal of Petroleum Exploration and Production Technology*, vol. 10, pp. 2871–2883, 2020.

[9] P. Zhang, C. Wang, and L. Li, "Deep transfer learning for seismic salt identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[10] TGS, "TGS Salt Identification Challenge," *Kaggle Competition*, 2018. [Online]. Available: https://www.kaggle.com/competitions/tgs-salt-identification

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3431–3440, 2015.

[12] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11, 2018.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and A. Rabinovich, "Going deeper with convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1–9, 2015.

[14] H. Di, L. Wu, and Y. Ma, "Seismic salt interpretation using deep convolutional neural networks," *Interpretation*, vol. 6, no. 3, pp. SE173–SE187, 2018.

[15] M. Zhou, J. Wu, and X. Liu, "Hybrid encoder-decoder networks with attention for salt body segmentation in seismic data," *IEEE Access*, vol. 10, pp. 11435–11446, 2022.