Deep Learning for Segmentation of Cement Spots in Core Images

Ivan Bolotov [1], Ekaterina Tolstaya [1], Mokhles Mezghani [2] and Aqeel Khalifa [2]

[1] *Aramco Research Center — Moscow, Aramco Innovations LLC, Moscow 119234, Russian Federation. E-mail: ivan.bolotov@aramcoinnovations.com (corresponding author); ekaterina.tolstaya@ aramcoinnovations.com*
[2] *EXPEC Advanced research Center Saudi Aramco, Dhahran, Saudi Arabia. E-mail: mokhles.mezghani@sauptech.com; aqeel.khalifa@sauptech.com*

Abstract

Detecting anhydrite spots on core images is an important task in oil&gas industry due to the impact of anhydrite on overall reservoir quality and production potential of hydrocarbon reservoirs. Accurate identification of anhydrite spots helps to differentiate zones of high and low permeability and identify areas with reduced reservoir quality. Traditional segmentation algorithms are often not accurate for anhydrite segmentation. Therefore, deep learning methods are promising to obtain stable and reliable segmentation results. However, the required training data are not always available. Within the presented research, we suggested utilizing synthetic data and preprocessing techniques to extend the training data for the task at hand. Our methodology for anhydrite segmentation relying on state-of-the-art neural networks showcased the ability to detect and delineate anhydrite spots of varying shapes and sizes on example of representative experimental data. Integrating proposed method into a comprehensive reservoir analysis could enhance the pipeline of geological analysis.

Keywords

Deep learning, segmentation, anhydrite, DCRF, Unet++, FPN, DeepLabV3, StyleGAN.

1. Introduction

Segmenting anhydrite spots in core images is important for several reasons, especially when the host rock contains some amount of anhydrite. Anhydrite ($CaSO_4$) was recently recognized [1] as a common mineral that can be found in sedimentary rocks and can significantly influence the rock's petrophysical properties, fluid flow characteristics, and overall reservoir quality. Identifying and quantifying anhydrite in core images can provide valuable insights into the reservoir's behavior and potential for hydrocarbon production. Anhydrite is generally less porous and permeable than the surrounding host rock. Accurate identification of anhydrite spots helps to differentiate between high and low permeability zones within the reservoir. This information is critical for reservoir modeling and production prediction. Anhydrite can act as a cementing agent, binding together other rock components, and reducing the overall reservoir quality. Understanding the distribution of anhydrite can help identify areas with reduced reservoir quality and potential drilling or production challenges. Anhydrite can form barriers to fluid flow within the reservoir. Identifying these barriers is essential for understanding fluid migration pathways, pressure communication, and potential compartmentalization within the reservoir. The presence of anhydrite can significantly affect the rock's mechanical properties, such as strength and brittleness. Knowledge of anhydrite distribution is important for wellbore stability, hydraulic fracturing design, and overall field development planning. Anhydrite can undergo mineralogical reactions under certain conditions, such as contact with formation water. These reactions can lead to the formation of other minerals like gypsum or calcite, which can impact reservoir properties and fluid flow characteristics. Understanding the distribution of anhydrite in the reservoir is crucial for designing and

GraphiCon 2023: 33rd International Conference on Computer Graphics and Vision, September 19-21, 2023
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
EMAIL: ivan.bolotov@aramcoinnovations.com (I. Bolotov); ekaterina.tolstaya@aramcoinnovations.com (E. Tolstaya); mokhles.mezghani@sauptech.com (M. Mezghani); aqeel.khalifa@sauptech.com (A.Khalifa)
ORCID: 0009-0007-8274-4773 (I. Bolotov); 0000-0002-8893-2683 (E. Tolstaya); 0000-0002-4843-5658 (M. Mezghani); 0000-0002-7843-1618 (A.Khalifa)

implementing enhanced oil recovery (EOR) techniques, such as waterflooding or CO_2 injection, as the presence of anhydrite can influence injectivity and sweep efficiency.

The goal of this work is to develop a methodology using modern technologies for segmenting anhydrite spots in core images. An example of a core image and manually segmented mask is shown in figure 1.

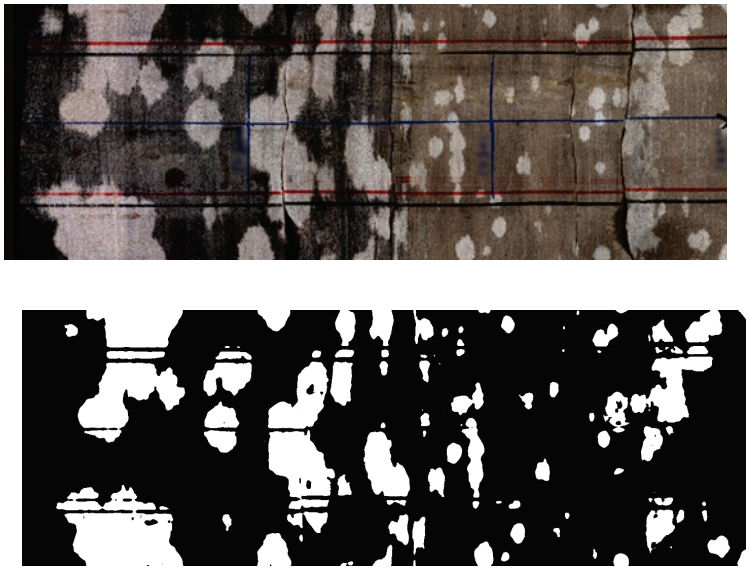

Figure 1: Example of core image and segmentation mask

The brighter spots can be found on darker rocks (background). In case of contrastive color, the spots can be segmented using simple thresholding, or Otsu thresholding [2]. As shown in figure 2, a brighter background is also possible.

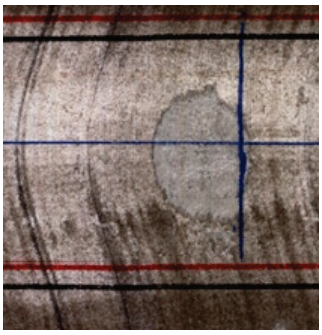

Figure 2: Example of a brighter background

In this case simple thresholding will not work. However, the spots have definite color palette, and can be easily visible on the images. Therefore, automatic method can be developed to address the problem. However, during the development of an automatic method for segmenting anhydrite spots, we encountered a problem of insufficient labeled data. This limitation hinders the development of accurate and reliable segmentation methods based on deep learning.

To address this issue, we propose utilizing synthetic data generated through two different methods: the first method involves combining existing samples in various ways, thus creating previously unseen images, while the second method is based on the application of StyleGAN. This approach allows us to expand the training dataset and provide diversity in conditions and contexts for model training.

The proposed methodology, based on deep learning, holds potential for stable and reliable segmentation of anhydrite spots. A comparative analysis of neural networks will enable us to select the optimal architecture for this task.

The implementation of the proposed methodology will have a significant impact on the oil industry. More precise and automated segmentation of anhydrite spots in core images will enable a more accurate assessment of reservoir properties, permeability, quality, and hydrocarbon production potential.

Consequently, this will lead to improved planning and optimization of extraction, enhanced efficiency in reservoir utilization, and improved production forecasting.

2. “Classical” image segmentation methods

2.1. Otsu thresholding

Otsu's thresholding is one of the methods for image binarization, which allows for automatic selection of an optimal threshold to separate an image into two classes: foreground and background. The Otsu's algorithm is based on minimizing the intra-class variance, which represents the spread of pixel values within each class. It searches for a threshold at which the variance within classes is minimized, while the difference between classes is maximized [2]. The Otsu's algorithm enables automatic selection of a threshold value for image binarization without the need for manual threshold setting. This makes it useful for various image processing tasks, including segmentation.

2.2. Dense CRF

DCRF (Dense Conditional Random Fields) refines segmentation output by considering pixel labels, surroundings, and spatial relationships. It aims to enhance pixel labels based on neighboring information, incorporating context and object connectivity. DCRF uses a Gibbs energy function to minimize label compatibility. The energy function includes unary potentials (independent label probabilities) and pairwise potentials (joint label probabilities for pixels). Pairwise potentials foster local object connectivity and segmentation consistency. "Dense" means all pixel pairs are considered. For multi-class segmentation, contrast-sensitive two-kernel potentials are used, considering color and position. DCRF improves object segmentation by leveraging label information, surroundings, and spatial locations [3].

In summary, DCRF is a method that refines the results of segmentation by considering information about pixel labels, their surroundings, and spatial locations. This enables obtaining more accurate and consistent object segmentation in an image.

2.3. K-means clustering

K-means clustering is a machine learning method used for grouping or clustering a dataset. In the context of image segmentation, each pixel is considered as a 3D vector, where each component represents the intensity or color channel of the pixel. The K-means clustering process begins by selecting a predefined number of clusters, K. Then, K initial cluster centers are randomly chosen. Each pixel in the image is assigned to the nearest cluster center based on the Euclidean distance in the vector space. Next, the center of each cluster is recalculated by computing the mean value of all vectors belonging to that cluster. This step is repeated until the cluster centers stabilize or a specified number of iterations is reached. The K-means clustering process aims to minimize the mean squared distance between each pixel and the center of its cluster. This allows pixels with similar intensity or color values to be grouped together in the same cluster. As a result of K-means clustering, each pixel is assigned a label corresponding to its cluster [4]. This enables the segmentation of the image into regions of similar pixels, which can represent object segments or background areas. K-means clustering is simple to implement and computationally efficient. However, it has certain limitations such as dependence on the initial choice of cluster centers and the possibility of converging to a local optimum. To achieve better results, it may be necessary to run the algorithm multiple times with different initial values.

2.4. SVM classifier

The Support Vector Machine (SVM) classifier is a machine learning method for image classification tasks. It separates data points into classes using a hyperplane, maximizing distance between them. SVM training optimizes a loss function to minimize errors and maximize class separation. Trained SVM

predicts class labels for image pixels. SVM can be trained on features like color spaces, and ensembles of SVM classifiers can enhance accuracy. SVM requires manual annotation for training data and sensitive parameter choices like kernel or distance function.

2.5. Baseline model results

In figure 3, the outcomes of classical algorithms are demonstrated. These outcomes can serve as a foundation for advanced deep-learning-driven image segmentation. This modern approach offers enhanced adaptability and stability across diverse images, particularly when aiming for heightened segmentation precision. Furthermore, it's conceivable to employ cascades of these techniques. For instance, the DCF technique could refine SVM classifier outputs or enhance results from clustering methods.

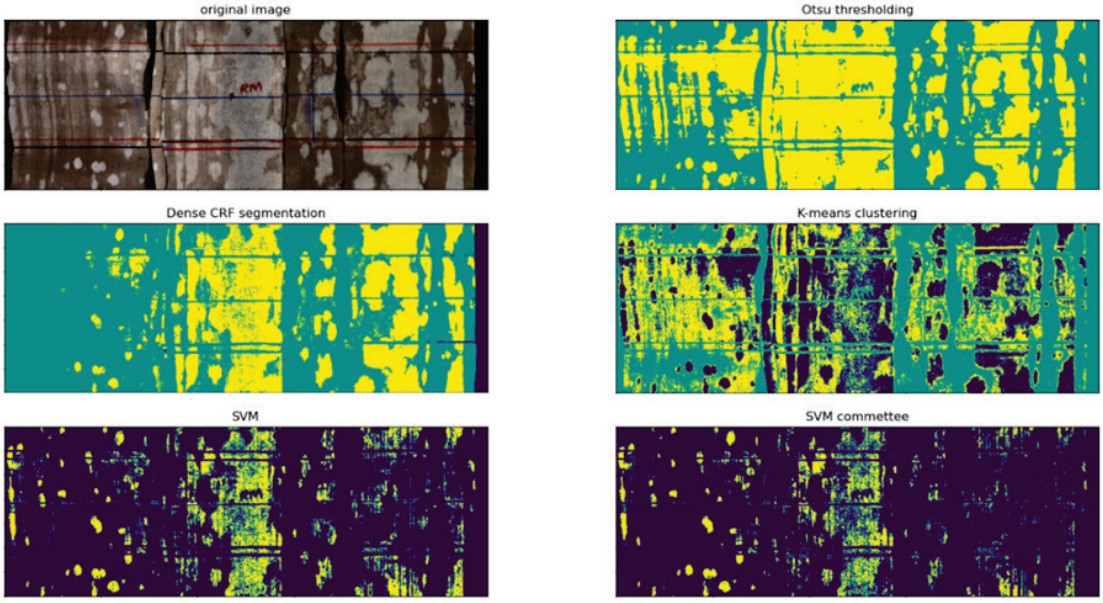

Figure 3: Example of segmentation using classical methods

Classical methods exhibit limitations in sustaining stable performance due to their inability to handle intricate and varied image patterns. These approaches often rely on manual feature engineering, rendering them less flexible when confronted with data variations. Complex image relationships and contextual nuances are frequently overlooked by classical algorithms, causing their efficacy to wane in the presence of intricate object boundaries, changing illumination conditions, and diverse textures. Therefore, in this study, image segmentation methods based on modern deep learning techniques will be employed. This will enable the model to adapt to various variations and contrasts of anhydrite patches. Such adaptability will allow maintaining stability and effectiveness of the model across different scenarios.

3. Neural networks-based image segmentation

Since classical approaches do not provide sufficient quality of segmentation masks, deep learning methods based on various neural network architectures have been developed. In this study, we will consider three architectures: U-Net, FPN, DeepLabV3+. This choice is motivated by the fact that these architectures are widely employed in related domains where similar tasks of segmentation and image analysis are addressed.

U-Net, with its ability to delineate intricate object boundaries, finds application in the medical field for detecting structures in organ images [5]. FPN, in turn, is often used for detecting objects of various sizes, which is crucial in satellite image processing [6]. DeepLabV3+ is renowned for its precise semantic segmentation capabilities, finding use in object recognition on streets.

The study compares these architectures because their successful application in diverse adjacent domains speaks to their universality and adaptability for various segmentation tasks. Through analyzing and comparing their performance in the context of segmenting objects with different shapes, sizes, and lighting conditions, we aim to determine which architecture might best suit our specific task of anhydrite patch segmentation.

3.1. U-Net

U-Net is a specialized deep learning model for semantic segmentation, classifying image pixels into categories. It was introduced by Ronneberger et al (2015) and adapted across domains. It features an encoding path (convolutions and max-pooling) and decoding path (up-convolutions and concatenations) forming a "U" shape, capturing local and global details (figure 4).

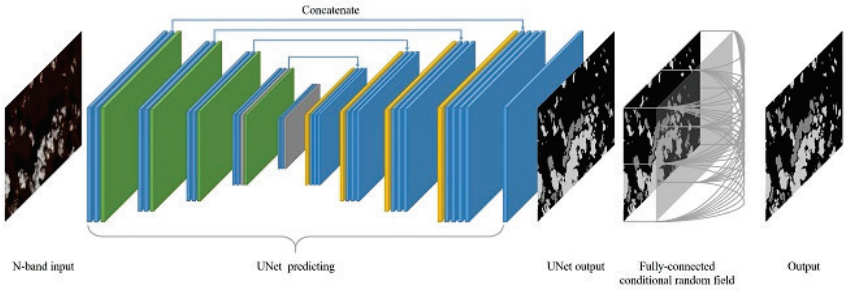

Figure 4: U-Net architecture

Encoding Path: Convolutions and max-pooling capture local features and expand receptive fields. Batch normalization and ReLU activation introduce non-linearity.

Decoding Path: Up-convolutions and concatenation restore spatial dimensions and incorporate high-res, local context.

Final Layer: A 1x1 convolution with softmax produces class probabilities, yielding the segmented output with input dimensions [5].

U-Net++ enhances the U-Net model by introducing a hierarchical decoder structure with multiple symmetric decoders at varying scales. This architecture aims to bridge the semantic gap between encoder and decoder feature maps prior to fusion. Each decoder receives and passes information between corresponding encoders, enabling efficient utilization of diverse feature levels for precise segmentation results [7].

U-Net++ achieves better information propagation and enhanced segmentation accuracy compared to U-Net (figure 5). Its strength lies in handling limited labeled data effectively. The architecture efficiently combines local and global context, enabling the learning of complex patterns, making it well-suited for various segmentation tasks like medical imaging, remote sensing, and scene segmentation.

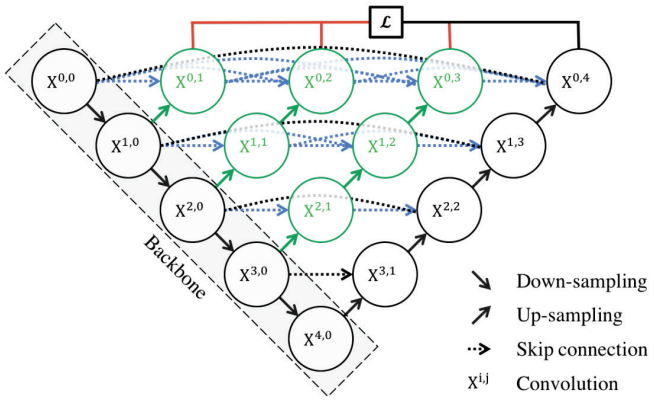

Figure 5: U-Net++ architecture [7]

3.2. Feature Pyramid Network (FPN)

The Feature Pyramid Network (FPN) architecture efficiently detects objects of different sizes. It addresses the loss of spatial information in CNNs due to pooling or deconvolution. FPN uses a feature pyramid from multiple encoder levels to combine context and object scale information (figure 6).

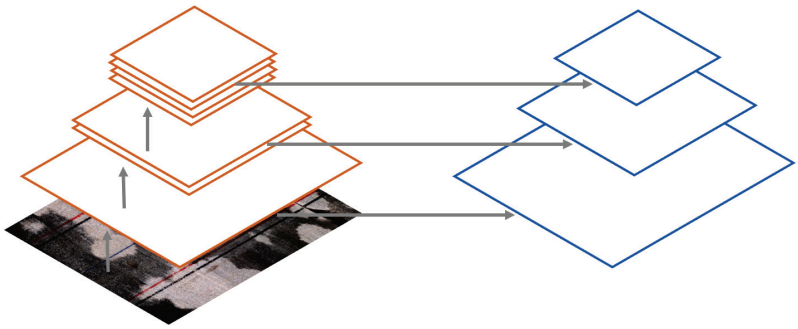

Figure 6: FPN architecture

Components of FPN:

1. Backbone Encoder: Utilizes networks like ResNet or VGG to create a feature pyramid, each level representing distinct object scales.
2. Feature Maps: Generated at each pyramid level with higher levels having semantic features and lower levels containing local details.
3. Feature Fusion: Combines semantic and local details via operations like summation or concatenation.
4. Upsampling: Upsampling module restores fused feature spatial resolution.

FPN effectively handles diverse object sizes by leveraging both high-level contextual and low-level detailed features. It's commonly used in object detection and segmentation for images with varying object scales [8].

3.3. DeepLabV3+

DeepLabV3+ is an advanced semantic image segmentation architecture. It builds on the DeepLab model family, known for state-of-the-art performance in pixel-level semantic segmentation tasks. Its main goal is effective capturing of local and global contextual information for accurate object segmentation. This is achieved through the encoder and decoder network combination, as shown in figure 7.

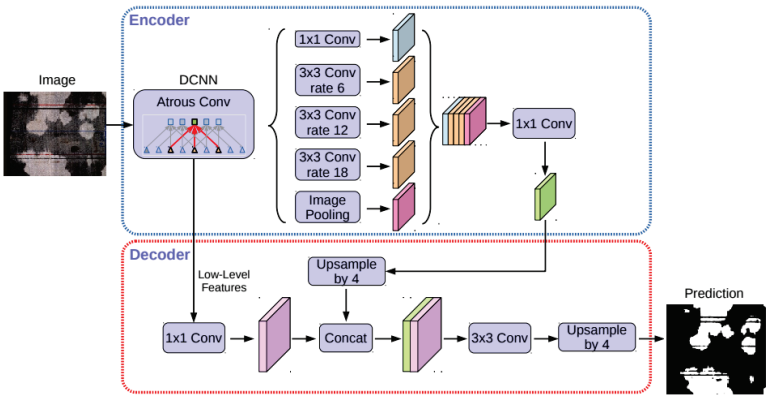

Figure 7: DeepLabV3+ architecture

The encoder network extracts high-level features from input images using pre-trained convolutional neural networks like ResNet or Xception. Atrous convolution, also known as dilated convolution, is integrated to increase receptive field without compromising spatial resolution. Atrous Spatial Pyramid Pooling (ASPP) further enhances multi-scale contextual understanding through parallel atrous convolutions with different dilation rates. The decoder network refines segmentation by using upsampling layers and skip connections, connecting low-level features from the encoder to incorporate detailed and semantic information [9].

Overall, DeepLabV3+ employs dilated convolutions, ASPP, and skip connections to capture local and global context, resulting in impressive semantic segmentation performance across various tasks.

4. Dataset

Initially, the dataset consisted of 19 images, each measuring 1620 by 4584 pixels. Almost all images contained grayish spots of various shapes and sizes. Manual annotation of the first 10 images proved to be challenging, so for the initial annotation stage, we employed the state-of-the-art "segment anything" model [10]. Currently, the model can take three arguments for image segmentation. The first argument is one or several points. In the case of multiple points, both the object and the background are marked. The second argument is the Bounding Box. The third argument is an already segmented mask. Typically, a mask obtained using the first two methods is input to improve segmentation. In our case, we utilized the Bounding Box approach, as illustrated in figure 8.

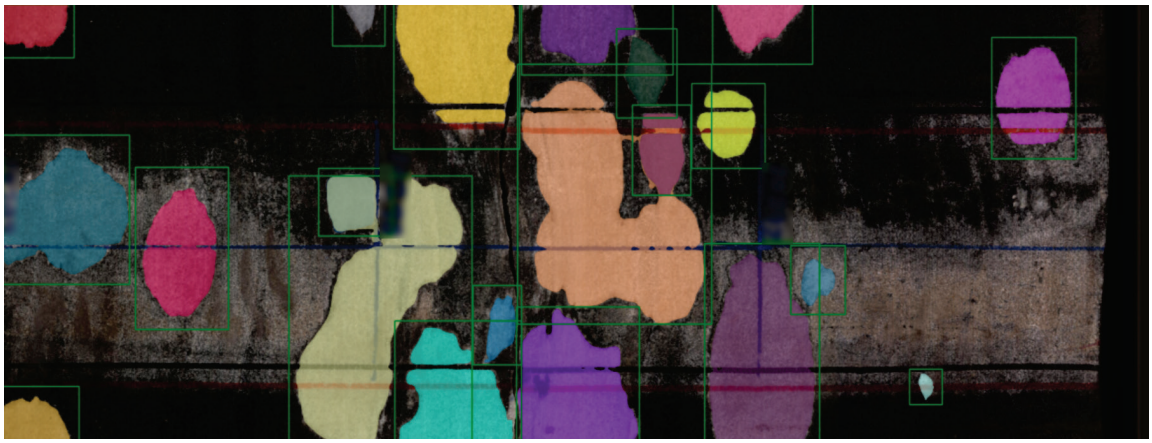

Figure 8: Annotating an image using SAM.

Subsequently, an initial version of the U-Net model was trained on the available masks. This model was applied to the remaining images, and following manual mask corrections, the new data was added to the training set. This dataset was validated by geologists. The second dataset contained about 150 images, almost each with dimensions of 2000 by 687 pixels. After applying the model to this new data, the masks were once again corrected and used as a new training dataset. The first dataset was retained as the validation dataset. Additionally, the dataset was expanded with synthetically generated samples. Synthetic data has emerged as a valuable resource for image segmentation tasks, providing an effective solution to the challenges posed by the scarcity of labeled data and the high costs associated with manual annotation. Synthetic data is artificially generated images with perfect ground truth labels. The use of synthetic data for training deep learning models in image segmentation tasks has proven to be beneficial in several ways. For synthetic samples generation we used two different techniques: first one based on combining in several ways the existing samples, thus creating never seen before images, and the second technique was in application of StyleGAN.

Existing samples combination includes overlaying segmentation spots from one image onto another. With the set of images with no spots or with small spots we generated combined images, taking spots with segmentation masks from images, where the spots occupy significant area of the image. We also applied augmentation prior to image overlay, so this slightly changed the geometry of the existing spots. The example of such image is presented in figure 9. StyleGAN, short for Style Generative Adversarial Network, is a state-of-the-art deep learning architecture introduced researcher Karras et al in 2019 [11].

StyleGAN builds upon the traditional GAN framework, which consists of a generator and a discriminator working in competition, to generate highly realistic and high-resolution images. The style-based generator in StyleGAN introduces a mapping network that converts a latent code into a style vector, which is then used to modulate the features at different layers of the synthesis network through adaptive instance normalization layers [11]. We applied conditional generation of images, inputting a binary mask, and outputting a synthetic core image with corresponding anhydrite spots (figure 10).

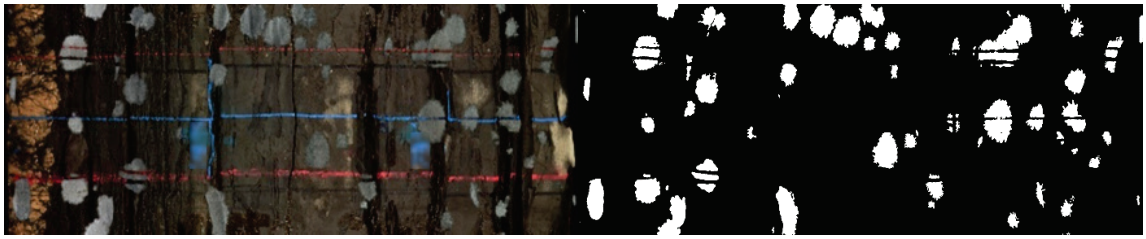

Figure 9: Example of synthetically generated image with corresponding mask, by combination of existing true images

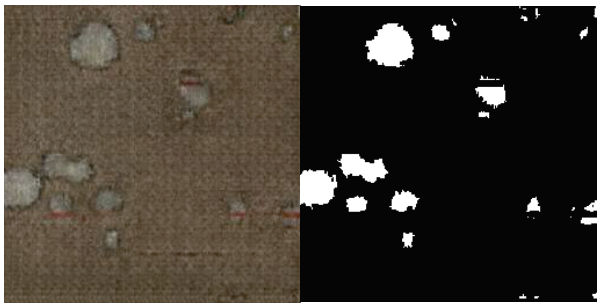

Figure 10: Examples of images, generated by StyleGAN

5. Training and evaluation

Image preprocessing was applied to manual masks to exclude colored lines that appear on the images of cores. This improved segmentation results as no colored parts are segmented on validation images. The lines are segmented out using saturation channel thresholding.

The model was trained on real and synthetic data. Patches of size 128 by 128 were extracted from high-resolution images and fed into the net, after augmentation step, which includes random shift, rotate, color modification, and flip. Loss function consists of weighted dice and cross-entropy losses.

6. Metrics

The model is evaluated using the following metrics: mean accuracy, mean intersection-over-union (*IoU*), and mean F1-score, using the first dataset, which did not take part in the training process, and was confirmed by geologists. In terms of true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*TN*), we computed the following metrics:

$$Accuracy: A = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$Intersection\text{-}over\text{-}Union: IoU = \frac{TP}{TP + FP + FN}, \quad (2)$$

$$F1\ score: F1 = \frac{2TP}{2TP + FP + FN}, \quad (3)$$

Accuracy: This metric (equation 1) measures the overall accuracy of the model by calculating the ratio of correctly classified pixels ($TP + TN$) to the total number of pixels ($TP + TN + FP + FN$). A value of 1 indicates perfect accuracy, while a value of 0 indicates complete inaccuracy.

IoU: The IoU (equation 2) coefficient measures the similarity between predicted and ground truth labels by calculating the ratio of the intersection area to the union area of the labels. The IoU coefficient

is computed for each class, and then the average is taken across all classes. A value of 1 indicates perfect overlap, while a value of 0 indicates no overlap.

F1-score: The F1-score (equation 3) is the harmonic mean of precision and recall. It measures the balance between precision and recall, considering both false positives and false negatives. The F1-score is computed for each class, and then the average is taken across all classes. The F1-score ranges from 0 to 1, where 1 indicates perfect precision and recall.

Figure 11 presents results of segmentation using different architectures of encoders. We can see here that U-Net++ architecture provides the best results in terms of our metrics.

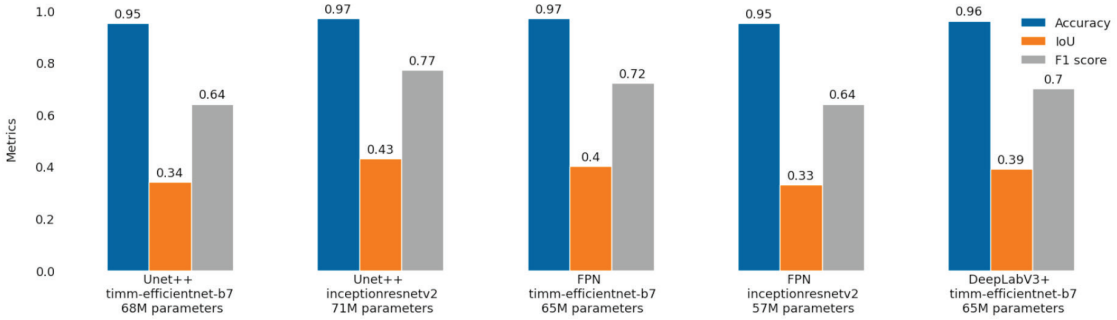

Figure 11: Comparative analysis of architectural segmentations

7. Postprocessing and results

Among the examined deep learning models, the U-Net++ architecture with an InceptionResNetV2 encoder demonstrated the best performance (figure 11). An example of a segmented mask is shown in figure 12. The figure illustrates that the obtained mask is not sufficiently accurate. Hence, the decision was made to employ the DCRF algorithm, outlined in Section 2.2, to enhance the output masks and metric results. Additionally, to increase mask prediction speed, input images were downsampled by half, which led to a substantial enhancement in the accuracy of the segmented mask. The masks were subsequently restored to the original resolution of the input images. The final post-processed result is presented in figure 13. Metric values underwent changes, as depicted in figure 14.

As a result of our research, a reliable methodology has been developed for the effective segmentation of anhydrite spots of various sizes in core images. This methodology encompasses the use of the state-of-the-art "segment anything" model for initial data annotation, as well as addressing the challenge of limited and unrepresentative datasets through the utilization of synthetic data, which generates novel and diverse, previously unseen, images.

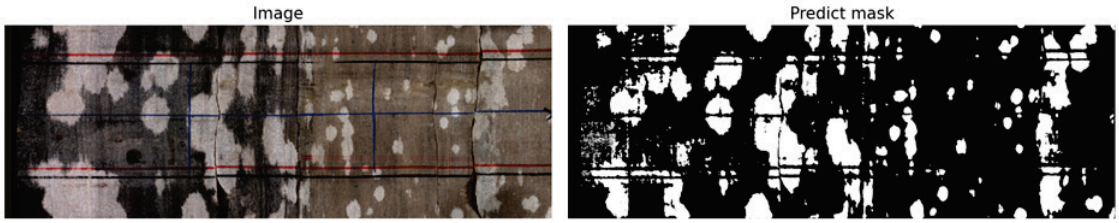

Figure 12: Image example and predicted mask

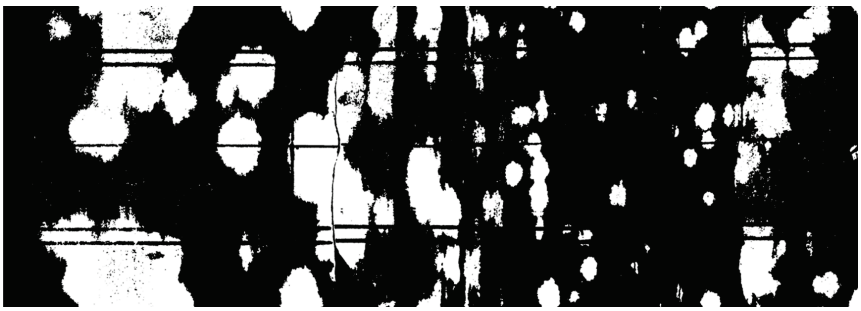

Figure 13: Predicted mask + DCRF

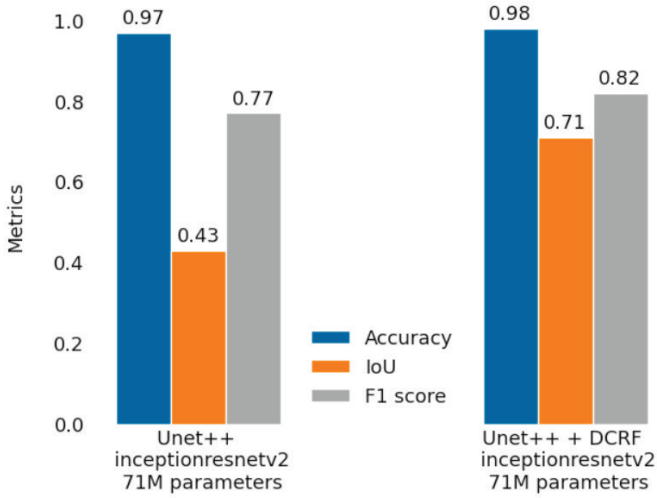

Figure 14: Changing metrics using the DCRF

8. Conclusion

This study successfully demonstrated the feasibility and efficacy of employing advanced image processing and deep learning techniques for the segmentation of anhydrite spots in core images. Through the utilization of a U-Net based architecture, optimized for the task at hand, a remarkable improvement in the segmentation accuracy and robustness compared to conventional methods was observed. Our methodology showcased the ability to detect and delineate anhydrite spots of varying shapes and sizes, which is of paramount importance in understanding their impact on reservoir properties. The precise segmentation of anhydrite spots facilitated a more accurate and comprehensive assessment of their distribution, which is integral to reservoir characterization, geomechanical analysis, and production optimization. By automating the segmentation process, not only is there a significant reduction in time and resources, but it also mitigates human-induced biases and errors, resulting in more consistent and reliable analysis. Moreover, the integration of synthetic data to augment the training dataset proved to be invaluable. This approach ensured a robust model capable of generalizing well to different scenarios, contributing to the model's versatility and applicability across diverse geological settings.

Future work could explore the incorporation of additional data sources and refine the network architecture to further improve the segmentation performance. Furthermore, integrating this segmentation framework into an end-to-end reservoir analysis pipeline could revolutionize the way geological core analysis is conducted. In summary, the developed segmentation methodology for anhydrite spots in core images is a powerful tool that has the potential to significantly advance reservoir characterization and management, paving the way for more informed and data-driven decision-making processes in the field of petroleum geology and reservoir engineering.

9. References

[1] Khalifa, A., Tarabbia, P., Alrowaie, M.A. and Soua, M., 2022. Significance and Implications of Diagenetic Nodular Anhydrite on Sandstone Reservoirs Quality. In SPE Annual Technical Conference and Exhibition. OnePetro. (pp. 1-10) doi:10.2118/210403-ms

[2] Yousefi, J., 2011. Image binarization using Otsu thresholding algorithm. Ontario, Canada: University of Guelph (pp. 1-4)

[3] Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with Gaussian Edge potentials advances in neural information processing systems, (pp. 109–117).

[4] Dhanachandra, N., Manglem, K. and Chanu, Y.J., 2015. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science, 54, (pp. 764-771). doi:10.1016/j.procs.2015.06.090

[5] Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241).
[6] Seferbekov, S., Iglovikov, V., Buslaev, A. and Shvets, A., 2018. Feature pyramid network for multi-class land segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 272-275). doi:10.1109/cvprw.2018.00051
[7] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., 2018. U-Net++: A nested U-Net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4 (pp. 3-11).
[8] Kirillov, A., Girshick, R., He, K. and Dollár, P., 2019. Panoptic feature pyramid networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 6399-6408). doi:10.1109/cvpr.2019.00656
[9] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818). doi:10.1007/978-3-030-01234-2_49
[10] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y. and Dollár, P., 2023. Segment anything. arXiv preprint arXiv:2304.02643 (pp. 1-5)
[11] Karras, T., Laine, S. and Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410). doi:10.1109/cvpr.2019.00453