Semantic rock-type segmentation of seismic images

Link for code:
https://drive.google.com/file/d/12MCXCSwvYplm6bNNxoOvfL0QhBkd7QTv/view?usp=sharing

Anshuman Pradhan, pradhan1@stanford.edu

Abstract

Building a model of the subsurface containing rock types from seismic data is a crucial task for petroleum industry applications. However, seismic data has challenges like lack of high frequencies which make predictions of rocks difficult. In this paper, we performed semantic segmentation of seismic images into rock types. We generated a synthetic training set for our network as seismic data is typically collected once in an oilfield. We designed and trained end-to-end fully convolutional networks with two different architectural units: encoder-decoder blocks and dilated convolutional layers. Preliminary analysis showed very good performance of both architectures in segmenting seismic images

Introduction

A critical research challenge for the geo-scientific community and petroleum industry is that of segmenting seismic images into classes of rock types and building a subsurface model. Such modeling can be crucial to making predictions about the performance of hydrocarbon reservoirs and deciding on optimum well-drilling locations. For instance, Figure 1 illustrates a synthetic subsurface depth section of channels filled with sand (yellow) in a background of shale rock (blue). The sandy channels provide optimum environment for generation and preservation of hydrocarbons and a desired goal in the industry is to classify rock-types and identify the spatial locations of these subsurface channels from seismic data acquired in the oil field. However, seismic waves typically have low frequencies and identification of thin (high frequencies) channels has been challenging. As can be seen in Figure 1, the seismic data is a smoothed, low frequency response of the rock patterns in the subsurface. It can be also be noted that thin channels, like the one highlighted in red in Figure 1, are below the seismic resolution and even lack discernible seismic signature. This limitation has traditionally impaired the performance of classification algorithms like logistic regression or discriminant analysis for this problem (Avseth et. al., 2001). This project aims to rethink solutions to this challenge, inspired by the recent advances made in semantic segmentation with deep convolutional networks. We expect the deep network to learn the highly non-linear mapping between the rocks and seismic images. The efficacy of different semantic segmentation architectures in segmenting the seismic image (input) into rock types (output) will be analyzed.

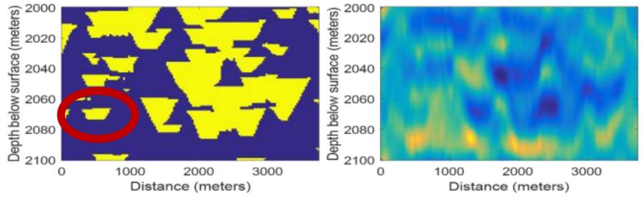

Figure 1: (Left) A synthetic subsurface depth section showing sandy channels (yellow) in a background of shale rock (blue). (Right) The corresponding modeled seismic section. Typically, seismic data is acquired in the field and the goal is to identify the spatial locations of these sandy channels.

Related Work

While conventional classification algorithms like discriminant analysis or Bayesian classification have been applied to pixel-wise classification of rock-types from seismic data (Avseth et. al., 2001, Avseth et. al., 2005), their applicability to thin reservoir intervals remains challenging due to the issues discussed previously. On the other hand, the deep learning community has witnessed extremely promising results in semantic segmentation of images. The tremendously rapid gains made in recent years by semantic segmentation applications were spurred by the seminal work of Long et. al., 2015. They used a fully convolutional network (FCN), trained end-to-end and pixel-to-pixel for semantic segmentation. Normally, the loss of resolution due to the pooling and subsampling layers is handled by transposed convolutional layers in the decoder network. Various kinds of decoder network architectures have been proposed in literature. Badrinarayanan et. al., 2015's encoder-decoder architecture has a hierarchy of decoder layers corresponding to each encoder layer. Upsampling is attained by unpooling with corresponding max-pooling layer indices. The loss of spatial resolution in the encoder part can sometimes be detrimental for dense predictions. Long et. al. 2015 address this issue by using skip connections which propagate fine scale information from lower layers. However, certain papers have deemed the usage of pooling/subsampling layers for dense prediction to be contentious. While these layers are effective in aggregating multi-scale information for global predictions, dense predictions demand a full resolution output. Methods have been proposed which attempt to make dense predictions without losing spatial resolution, while still gathering multi-scale information. Dilated convolutional layers are used for this purpose, which can exponentially increase the receptive field of the layers (Yu and Koltun, 2015). Chen et. al., 2016 also integrate fully connected conditional random fields with dilated convolutional networks for boosting the network's ability to capture fine scale details. We decided to analyze and compare the applicability of FCNs with encoder-decoder architecture and dilated convolutional architecture to seismic rock classification problems.

Synthesis of training data set

A point of critical importance is that seismic data is generally acquired once for an oil-field. In order to create multiple training examples for machine learning applications, a synthetic training set needs to be generated. This can be achieved by geostatistical simulations and subsequently seismic wave propagation modeling. For example, a geologist might encapsulate his prior beliefs about the distribution of rocks in a prior image as shown in Figure 2. The prior image is used to generate multiple realizations of possible subsurface depth sections by geo-statistical simulations (Mariethoz and Caers, 2014). For the purposes of this preliminary proof-of-concept project, we decided to use two rock-types: sand and shale. Our subsurface rock images contain patterns which are commonly encountered in deltaic depositional settings. The images are 120*150 and contain pixel-wise labels for the rock types. Seismic sections are simulated from the rock sections using seismic wave propagation physics. The seismic images are 120*150 floating-point matrices containing Born-filtered seismic wave amplitudes. Since the aim was to train moderately deep FCNs end-to-end, we decided to use a train/dev/test split of 25000/2500/2500. It should be noted that we expect the network to learn the inverse function mapping from seismic data to subsurface rock types. Once learned, this mapping can be used to predict rock types from the real data acquired in the field. The reader is referred to the "Future Work " section for a detailed discussion of the nuances and implications of real-data predictions by a network trained with synthetic data.

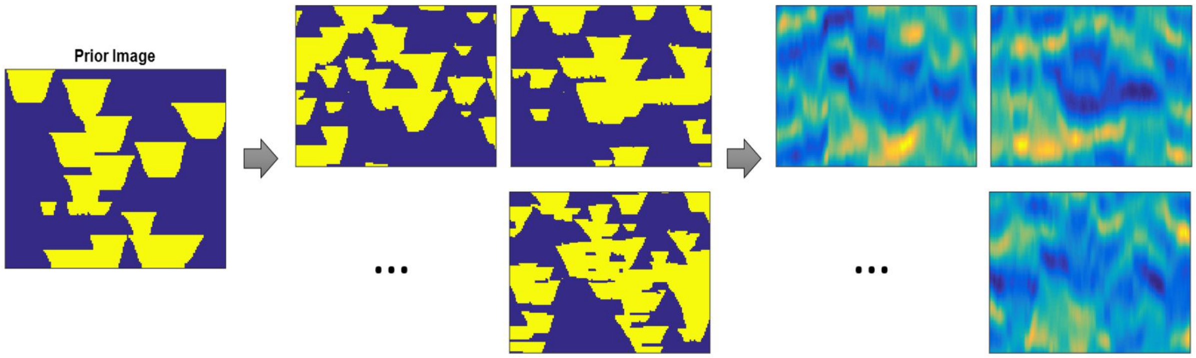

Figure 2: A prior image (left) is used to generate subsurface rock realizations by geo-statistical simulations. The corresponding seismic sections are modeled using seismic wave propagation physics

Network architectures

Two variants of semantic segmentation architectures were designed for end-to-end training:

- **FCN with encoder-decoder architecture (FCN-EncDec)**: The baseline convolutional network (FCN-EncDec-Baseline) architecture was inspired by the common encoder-decoder architectures employed for semantic segmentation applications (Badrinarayanan et. al., 2015, Long et. al., 2015). As shown in Figure 3, the encoder part of the network consists of two convolutional layers interspersed by max pooling layers. The number of layers was chosen to keep the baseline model as simplistic as possible. The convolutional layers use 3*3 filters with unit strides and 'SAME' padding. From encodings of size 30*37, we revert the effect of each convolutional and max pooling layers with transposed convolution. We also include RELU non-linearities in the decoder so as to enable the network to learn a non-linear upsampling. The output layer has a sigmoid activation function to predict the class labels. We chose not to use unpooling layers for upsampling as it was felt the learnable upsampling possible with transposed convolution layers offers more flexibility. A deeper network (FCN-EncDec-Deep) was also analyzed which contained two convolutional layers before each pooling layer in Figure 3 and corresponding transposed convolutional layers in the decoder network.

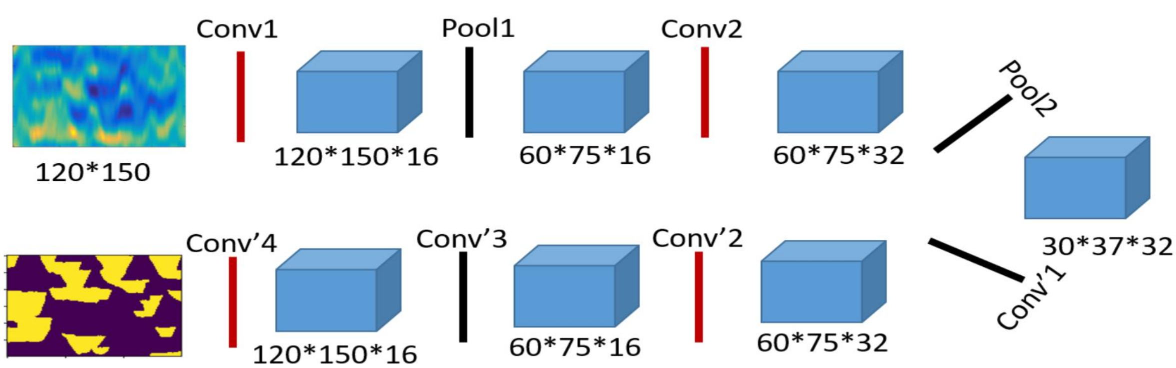

Figure 3: The encoder-decoder network architecture (FCN-EncDec-baseline). The inputs consist of seismic images and outputs are rock-types for each input pixel

- **FCN with dilated convolutional architecture (FCN-Dilated)**: Our encoder-decoder architecture made no use of methods like skip connections to incorporate fine scale information from shallower layers.

Skip connections require the storage and handling of activations from shallower layers, which increase the burden on computational requirements. We wanted to analyze whether the architectural stipulation of downsampling the input only to follow it up by upsampling is actually necessary for our application. We therefore designed an architecture with dilated convolutional layers (Yu and Koltun, 2015), which pad the filter with zeros, facilitating exponential increase in the receptive field of layers without the necessity of pooling. The network architecture is shown in Figure 4. The layers use 3*3 filters with unit strides and 'SAME' padding. Unit dilation factors were used for the first convolutional layer. Subsequently, the dilation factors were increased by a factor of 2, facilitating exponential increase in receptive field, till the receptive field of the layers covered the entire input image.

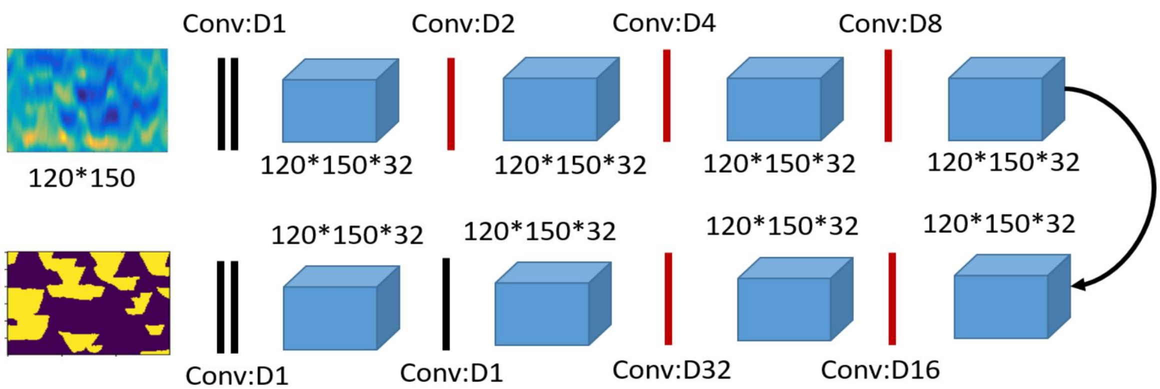

Figure 4: The dilated convolutional network architecture. The dilation factors are shown at the top of each layer

Training the network

We used the sigmoid cross-entropy loss summed over all the pixels in the image as the loss function. The performance analysis metric we employ is the mean classification accuracy at pixel. The network was trained with the Adam Optimization algorithm. An initial run of the base-line model training showed decent performance on dev and test sets (Table 1). In order to boost the network's performance, batch normalization layers were introduced after all convolutional layers. Batch-norm seemed to boost network performance significantly. Since the network, in general, showed good performance, the goal was to boost the performance as much as possible. Hence, we also used a deeper network with encoder-decoder architecture as discussed previously. Hyper-parameter search was performed over learning rate and mini-batch size. Learning parameter seemed to be most sensitive hyper-parameter. The results were not very sensitive to mini-batch size. The optimized hyper-parameters are shown in Table 1. No regularization was used since the model did not seem to be overfitting. The training was run for 100 epochs.

Table 1: Parameters and accuracies of networks trained for 100 epochs

Model	Batch-Norm	Learning rate	Batch-size	Train accuracy	Dev accuracy	Test accuracy
FCN-EncDec-Baseline	No	0.001	256	0.866	0.8772	0.877
FCN-EncDec-Deep	Yes	0.001905	256	0.946	0.9429	0.9427
FCN-Dilated	Yes	0.0079	64	0.953	0.9492	0.9499

Results

The best accuracies obtained on dev/test sets are shown in Table 1. We compare the predicted segmentations by FCN-Dilated against the ground truth for two examples from the test set. It can be observed that the trained network has done a good job at segmenting the seismic image. Especially, we found the network to have successfully segmented thin channels from seismic images (highlighted portions in top row of Figure 5). This is generally deemed to be challenging since seismic waves have lack very high frequencies. While the network has performed generally well, it has failed to capture the rock boundaries precisely (highlighted portions in bottom row of Figure 5). Observing the seismic data closely in Figure 1, we can see that seismic amplitudes normally lack resolution at the edges of the rocks. Better performance can be enforced at the edges by designing a loss function with high penalty at the edges. In terms of comparing the network architectures, the dilated convolutional architecture seemed to slightly improve performance. The encoder-decoder network produced similar results without enforcing any fine scale information from shallow layers.

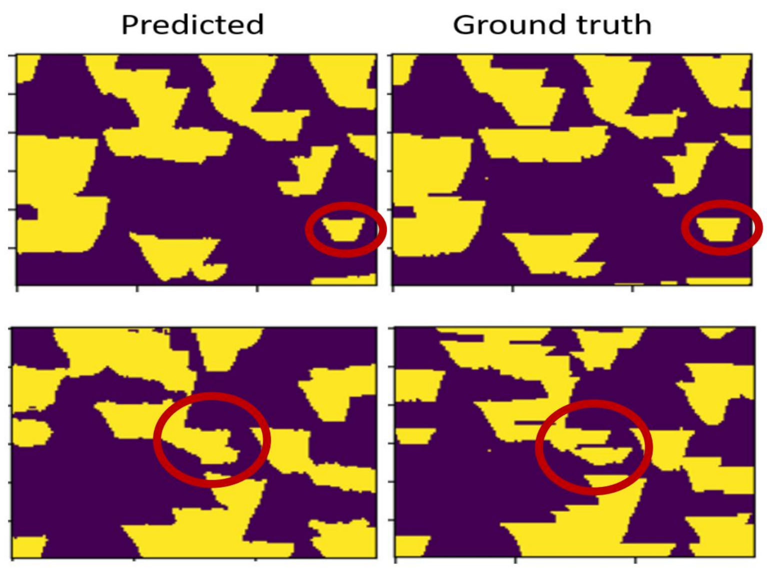

Figure 5: Predicted (left) segmentations by FCN-Dilated against the ground truth (right) for two test set examples

Discussion and Future Work

The efficacy of two semantic segmentation architectures for predicting pixel-wise rock types from seismic images was analyzed. Our preliminary results show that these architectures can be very efficient in segmenting seismic images. However, it should be noted that while our network was trained on a synthetic dataset, the actual predictions would be made on the real seismic data collected from the oil-field. Thus, the training set should ideally contain all the possible types and patterns (the sand rocks had channel-like patterns in our example) of rocks. In future, we plan to generate a more challenging training dataset from multiple prior images containing multiple patterns and types of rocks. Given that the initial iteration of the project was intended as proof of concept and the current lack of access to any actual field seismic data, our dev/test set contained examples from the same prior image. For prediction on real datasest, it should be made sure that the dev/test set contain examples which are specific to the oil field where this method is intended to be applied. We plan to adapt our network to predict rock types from real field dataset in the future.

References

Avseth, P., T. Mukerji, A. Jorstad, G. Mavko, and T. Veggeland, 2001, Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system: Geophysics, 66, 1157–1176.

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." arXiv preprint arXiv:1511.00561(2015).

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915 (2016)

Gregoire Mariethoz and Jef Caers. Multiple-point Geostatistics: Stochastic Modeling with Training Images. 2014. ISBN 9781118662953. doi: 10.1002/9781118662953.

Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: CVPR. volume 07-12-June; 2015. p. 3431–40. doi:10.1109/CVPR.2015.7298965. arXiv:1411.4038.

Yu, F., and Koltun, V., Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122, 2015.