Introduction

In salt provinces of the Gulf of Mexico (GOM), Angola and Brazil, the precise geometry of salt bodies is crucial for accurate seismic imaging, hydrocarbon reservoir exploration and planning of well paths. Therefore, salt interpretation has been a major part of the routine seismic imaging workflow. However, the precise identification of salt bodies generally takes about one third to half of the duration of a seismic data processing project because it depends mainly on human expertise of interpretation of salt bodies. Unfortunately, manual salt-picking is not only laborious and time-consuming, but also subject to human errors because of poor qualities of seismic images that have been contaminated with strong noise and amplitude variations. With the optimization of seismic processing cycle time and quality, the automatic salt body delineation becomes highly demanding.

Fortunately, salt interpretation can be considered as one of computer vision problems that might be resolved with deep learning. Inspired by TGS Salt Identification Challenge on a Kaggle platform (Kaggle, 2018; Karchevskiy et al., 2018; Babakhi et al., 2019), we have integrated several state-of-the-art techniques in deep learning to our solutions for accurately identifying salt bodies, which include, but are not limited to, UNet-like neural net with Imagenet-pretrained ResNet (He et al., 2016) as backbone, attention mechanism (Vaswani et al., 2017; Wang et al., 2017) for salt body focusing, one-cycle learning rate policy (Smith, 2018) for training acceleration, fast-SWA (Stochastic Weight Averaging) for better model generalization (Athiwaratkun et al., 2018), etc.

Network architecture

We follow a U-Net architecture (Ronneberger et al., 2015) to build our neural network for salt identification, dubbed accordingly U-SaltNet, which is composed of both a contracting encoder and an expansive decoder, as well as skip connections between blocks of encoder and decoder to recover original spatial resolutions. To improve prediction accuracy of salt bodies, we make several important network modifications. For example, a ScSE (concurrent Spatial Channel Squeeze and Excitation) module (Roy et al., 2018) are embedded in each block of both the encoder and decoder of our U-SaltNet to emphasize informative features and suppress irrelevant ones. In addition, we also incorporate a Feature Pyramid Attention (FPA) module (Li et al., 2018; Babakhi et al., 2019) to our U-SaltNet to increase the receptive field by fusing features from pyramid scales and provide precise pixel-level prediction. Figure 1 shows the network architectures.

Figure 1. neural network architecture of U-SaltNet. Note the encoder (Resnet-like structure) and decoder are not symmetrically implemented. For AG, please check Schlemper et al. (2018) for details.

To highlight salient regions in feature maps, improve model sensitivity and prediction accuracy, we further embed Attention Gates (AG) module (Schlemper et al., 2018) on skip connections right before the concatenation. This allows us to not only exploit the high-level feature map to guide low-level features and recover pixel localization but also focus on relevant activations. In this way, information extracted from a coarse scale is used in gating to reduce irrelevant and noisy responses from input feature
maps. Furthermore, AGs allow one to generate a fine-grained attention map that provides better insights into how model predictions are made and support potential interpretable deep learning. Finally, Hypercolumns (Hariharan et al., 2015) are also included in U-SaltNet to extract more information.

**Implementation**

The inputs are batches of 2D fixed-size tiles extracted from a 3D volume of depth-migrated seismic image. In addition to normal data augmentation techniques, Validation Time Augmentation (VTA) and Test Time Augmentation (TTA) with left-right flipping are also implemented in training and prediction phases, respectively.

We employed Lovasz hinge loss function (Berman et al., 2018) as our loss function in neural network due to its direct optimization of the Intersection Over Union (IoU). We train the network using a mini-batch Stochastic Gradient Descent (SGD) optimizer with a variant of a one-cycle learning rate policy (Smith, 2018). The learning rate at 0.004 turns out to be a good starting value. To further accelerate convergence and improve the generalization capacity of the networks, we have applied fast-SWA (Stochastic Weight Averaging), which averages the multiple weights along the trajectory of SGD within each cycle of a cyclical learning rate schedule (Athiwaratkun et al., 2018).

The Resnet-like encoder was initialized with ImageNet’s pre-trained model. The training epoch is set to 200, although the network typically converges in the first 10’s of epochs and no significant improvement has been observed beyond that. To further accelerate network training, we take advantage of the distributed synchronous SGD that distributes large mini-batches of data over a cluster system. We also utilize mixed-precision computing with GPUs, which saves memory and improves throughput of pure FP16 training while matching the accuracy of FP32 training.

We apply K-fold cross-validation to track the quality of the models and prevent overfitting. Every model is trained K times, i.e., one per fold, and the final score is the average or majority vote of the predictions of these models in the ensemble.

**Examples**

We evaluate our approach with two real data examples of 3D depth-imaged seismic volumes, GOM Area I and Area II, respectively. Besides seismic images, their corresponding salt masks, manually-picked by seismic interpreters, are also available and are used for the purpose of either training or validations. The goal is to delineate salt bodies from their sediment background.

![Figure 2](image-url)  
**Figure 2.** Cross validation of salt identifications from two separate GOM RTM-migrated images. The top row represents the results for Area I and the bottom row is for Area II. a) an inline slice of predicted salt mask overlaid on Area I image; b) manually-picked salt mask (the ground truth) by interpreters; while c) and d) are corresponding results for Area II image.
To leverage the existing network architecture pretrained with the ImageNet dataset, we have split the original 3D raw seismic image volume into many 2D tiles with fixed-sizes (128 or 256 samples). The resulting 2D tiles can be considered as a set of single-channel grayscale images with some of them containing salt bodies. These tiles are further split into K-folds (with K=5) with each fold consisting of two non-overlapping groups of tiles for the purpose of training and validations, respectively.

To test the generality of our U-SaltNet, we utilize both GOM images to cross validate each other. For example, we would predict salt mask of Area II image with the network model trained with Area I image and its manually-picked salt mask, and vice versa. The trainings generally took several hours on one node with 4 P-100 GPUs, while the predictions took only minutes for one line of images.

Figure 2 is our results of cross validation of Area I and Area II depth-migrated images, both of which are migrated with final velocity models. Note the predicted salt bodies (Figure 2a) with our algorithm match reasonably well with input seismic image and manual interpretation (Figure 2b). Similar results are achieved for Area II with the model trained with Area I Image and its salt mask (Figures 2c and 2d). Note the salt masks at the edges in Figure 2d are actually manually picked from a larger image, while we are testing a smaller size of the original image.

![Figure 2: Predicted and Manual Salt Masks](image)

**Figure 3.** Salt mask prediction from seismic image (Area I), migrated with salt flood model. a) predicted salt mask overlaid on seismic image; b) salt mask (the ground truth) manually-picked from final seismic image, same as Figure 2b. Notice the similarities between salt masks in a) and b).

Figure 3 compares the result of the predicted salt mask (Figure 3a) from an Area I image to the salt mask (Figure 3b) manually-picked from the Area I final image. The network model is trained with the given Area II’s image and salt mask. Note the two seismic images overlaid in Figure 3 are not the same. In specific, the seismic images in Figure 3a is RTM-migrated with a salt flood model, while the seismic image in Figure 3b is migrated with the final velocity model. Obviously, several iterations of salt model building are necessary to get to the final velocity model. But it demonstrates the potentials of our U-SaltNet for salt velocity model buildings.

The attention map for network qc is illustrated in Figure 4b, along with predicted (Figure 4c) and manually-picked salt bodies (Figure 4d) from a randomly-selected tile of the seismic image (Figure 4a). This was done by generating the gating signal to pinpoint local as well as global information that is highly beneficial for salt localization. This is especially beneficial for variable and small size salt bodies.

![Figure 4: Attention Map](image)

**Figure 4.** Attention map for a randomly selected tile of Area II image. a) a randomly-selected tile of seismic image; b) attention map to focus on salt body; c) predicted salt mask; d) manually-picked salt mask (the ground truth).
Conclusions

We have proposed a deep learning neural network architecture, which incorporates several existing state-of-the-art modules, for the task of salt body interpretation. These modules include, but not limited to, ScSE, FPA and AG modules. The experimental results of the proposed solution to real data examples indicate that it is capable of detecting subtle salt features with sufficient precision to be useful in seismic imaging and interpretation workflows.

Acknowledgements

We would like to thank Equinor for permission to present this abstract and Mike Cogan for his thorough reviews and revisions of the manuscript.

References


