SaltSeg: A β-variational autoencoder constrained encoder-decoder architecture for accurate geologic interpretation

Satyakee Sen, Sribharath Kainkaryam, Cen Ong, Arvind Sharma, TGS*

Summary

We describe SaltSeg: a high capacity deep convolutional neural network (CNN) architecture that achieves human level interpretation accuracy on seismic images. SaltSeg is primarily used for salt interpretation and is a key component of a deep learning based fully automated salt model building pipeline. It is designed to work on low resolution, noisy, incorrectly migrated seismic images as is typically encountered during the model building stage and achieves human level interpretation accuracy on such images and can be easily modified for other geologic interpretation tasks. We give an indepth description of the key building blocks of SaltSeg and describe a novel integration of a β-variational autoencoder (VAE) branch with a standard encoder-decoder network that leads to significant boost in interpretation accuracy. We validate our results using real data images from surveys in the Gulf of Mexico.

Introduction

The architecture design of a convolutional neural network (CNN) plays a key role in the success of the segmentation or classification task that needs to be performed. Seismic images are inherently different from natural images (band-limited nature and high noise level) and thus straightforward adaptations of CNN architectures usually leads to models that do not have the desired level of accuracy for successful production deployment. In this paper we describe a semantic segmentation network: *SaltSeg*, custom designed to work on seismic images.

Architecture design

SaltSeg's base architecture follows a conventional encoder-decoder (Ronneberger et al., 2015) set up where the encoder blocks are used to efficiently encode the key features of the input image, while the decoder blocks tries to to reconstruct a semantic segmentation map of the input image constrained by the encoder blocks. On top of this base architecture we add hypercolumns (Hariharan et al., 2015) and a β-VAE branch (Higgins et al., 2017). In the following sections we describe in detail the key components of SaltSeg and the accuracy improvements achieved by using design components specific to seismic data. We use salt interpretation during the salt model building flow as our test case to demonstrate the robustness of SaltSeg. It is well known that migrated volumes during the salt model building stage are prone to be contaminated by strong migration artefacts, have low resolution and the imaging is often unclear or ambiguous and thus salt picking on such images can serve as a good yardstick to gauge the accuracy of CNN based geologic interpretation. We will show that SaltSeg achieves human level interpretation accuracy on such images. Further all results shown in the next few sections are *blind test* results , i.e., SaltSeg was trained on legacy surveys and applied on images from new surveys, no part of which was in the training set nor did the new survey have any overlap with the legacy surveys used to build the trining set.

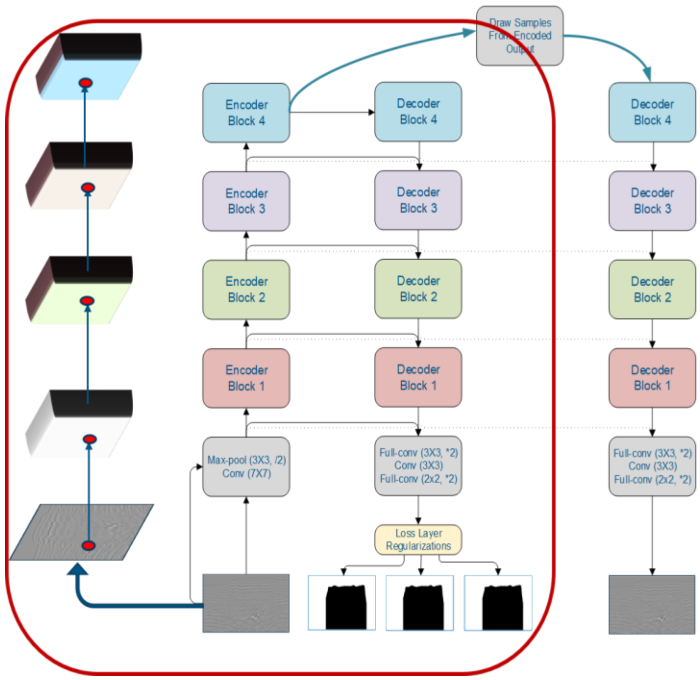
Encoder blocks

The encoder part of SaltSeg is adapted from a powerful architecture used in biomedical image segmentation, LinkNet (Chaurasia et al, 2017). In contrast to the original LinkNet, which uses Resnet18 (He et al., 2016) as the encoder, we use a more powerful 34-layer deep Resnet34 architecture. Our motivation to switch to a deeper and higher capacity archictecture stems from the complexity of the underlying segmentation task (working with noisy seismic images). The encoder reads in large image patches of size (384 X 384) and progressively downsamples the spatial dimensions while adding feature maps to efficiently encode the image's key features. The top blue table in Figure 1 shows the details of each encoder block. Note that the final encoder block uses 512 feature maps to encode the original image. Figure 4 shows the inaccuracies in the raw probabilities predicted by a shallower (Figure 4b) version of our network compared to our deep architecture (Figure 4c) for picking the salt flood mask on a complex sediment flood migrated image (Figure 2a). We note that a well designed deeper encoder network would almost always outperform its shallower counterpart especially when the task is complex.

Decoder blocks

SaltSeg has four decoder blocks maintaining symmetry with the encoder blocks used on the encoder side. The construction of each decoder block is identical and described in the left blue table in Figure 1. Unlike a conventional U-Net or encoder-decoder architecture each decoder block receives input from the corresponding encoder block as well as the encoder block below it (curved straight lines in red-box of Figure 1). This imposes additional constraints on the decoder's upsampling operation.

10.1190/segam2019-3216875.1



Encoder block #	Layers	Channels	Output image size
Encoder 1	6	64	96 X 96
Encoder 2	8	128	48 X 48
Encoder 3	12	256	24 X 24
Encoder 4	6	512	12 X 12

Decoder block i operations
Add encoder-i output
Nearest neighbor interpolation
Conv 3X3
Add encoder-(i-1) output

VAE Draw samples block
ReLU
Conv (8) stride 2
2 FC layer (512)
Draw mean (256X1) and std (256X1)

Figure 1: SaltSeg architecture: Red box is a standard Resne34 based encoder-decoder The leftmost branch is a hypercolumn branch that tracks spatial location of each pixel in the input image. The loss-layer in the standard encoder-decoder setup can be trained with various high entropy promoting regularizations. The right most branch is the β-VAE branch which draws samples from the final encoder layer to estimate high dimensional mean and vaiance and uses this to reconstruct the input image fully. The tables indicate (top) encoder layer description, (left) units in each decoder block all of which are identical and (right) the operations performed to draw a sample mean and variance for the VAE branch. Conv:convolution, FC: fully connected.

β-VAE Branch

The key design aspect of SaltSeg is the use of a β-VAE branch (the rightmost branch in Figure 1). The purpose of this branch is to reconstruct the input image completely, i.e. it aims to learn the identity function. By forcing SaltSeg to perform this auxillary task, during training, of trying to reconstruct the image, the encoder is contrained to learn the most relevant features of the image such that the Identity mapping by the VAE branch can be done. Since the segmentation-decoder branch shares the same encoder block, better features learned by the encoder means better segmentation performance. The loss function used by SaltSeg is a weighted combination of the losses coming from the segmentation branch and the VAE branch, given by:

$$\boldsymbol{Loss = w_1(BCE + 1 - DICE) + w_2 L_2 + \beta L_{KL}}$$

The first two terms are the standard image segmentation losses coming from the segmentation branch which combines binary cross entropy (*BCE*) with the differentiable *DICE* score (Milletari et al., 2016). The last two terms are the losses from the VAE branch. The L_2 loss is the pixel wise mean squared error between the input image and the reconstructed image, while L_{KL} measures the distance between the distribution of the reconstruction and the input data via the Kullback-Liebler (KL) divergence. The KL divergence is weighted by a factor β>1 which encourages the VAE branch to prefer encoder reprentations that show greater disentaglement (less overlap in the feature maps), i.e. the encoder is forced to learn feature maps which are more diverse.

Hypercolumns

It is well known that as the images flows through the network, the spatial information is lost while semantic information is enriched via the increasing number of feature maps. However for complex, noisy images like those encountered in seismic imaging, we track the explicit spatial location of each pixel in the input image as it flows through the network using hypercolumns, which are a tall column vector (one for each sample in the image) that records the activations in the layers of the network above

10.1190/segam2019-3216875.1

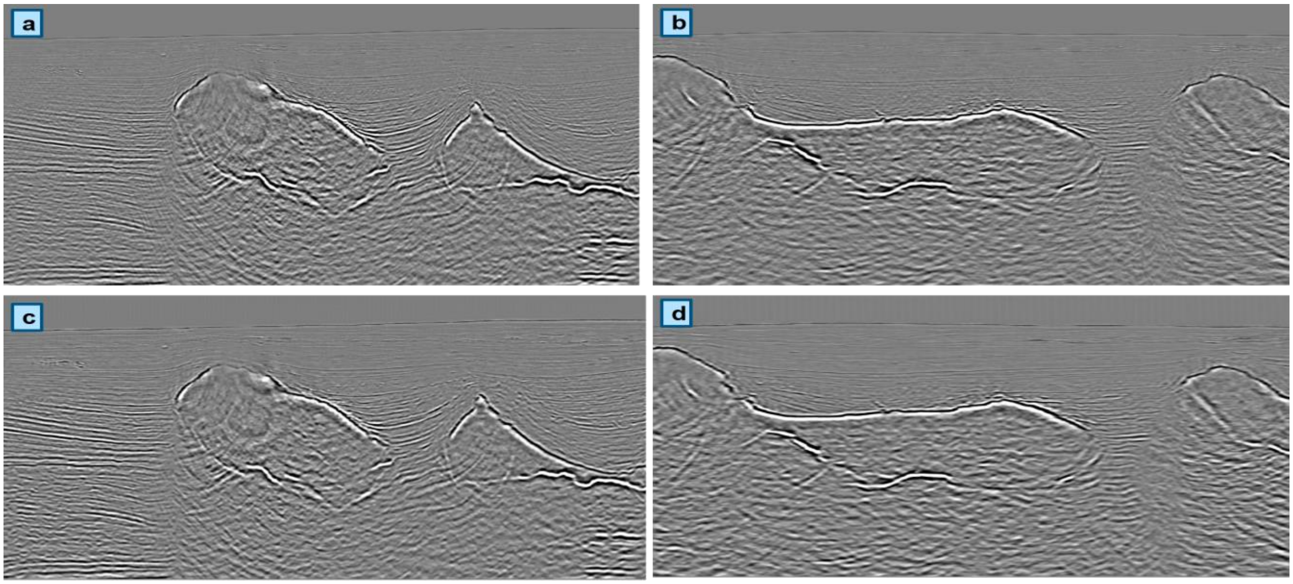


Figure 2: Reconstruction capability of the VAE branch on salt flood migrated images from new survey never seen during training (a) original inline, (b) original crossline, VAE reconstructed (c) inline and (d) crossline.

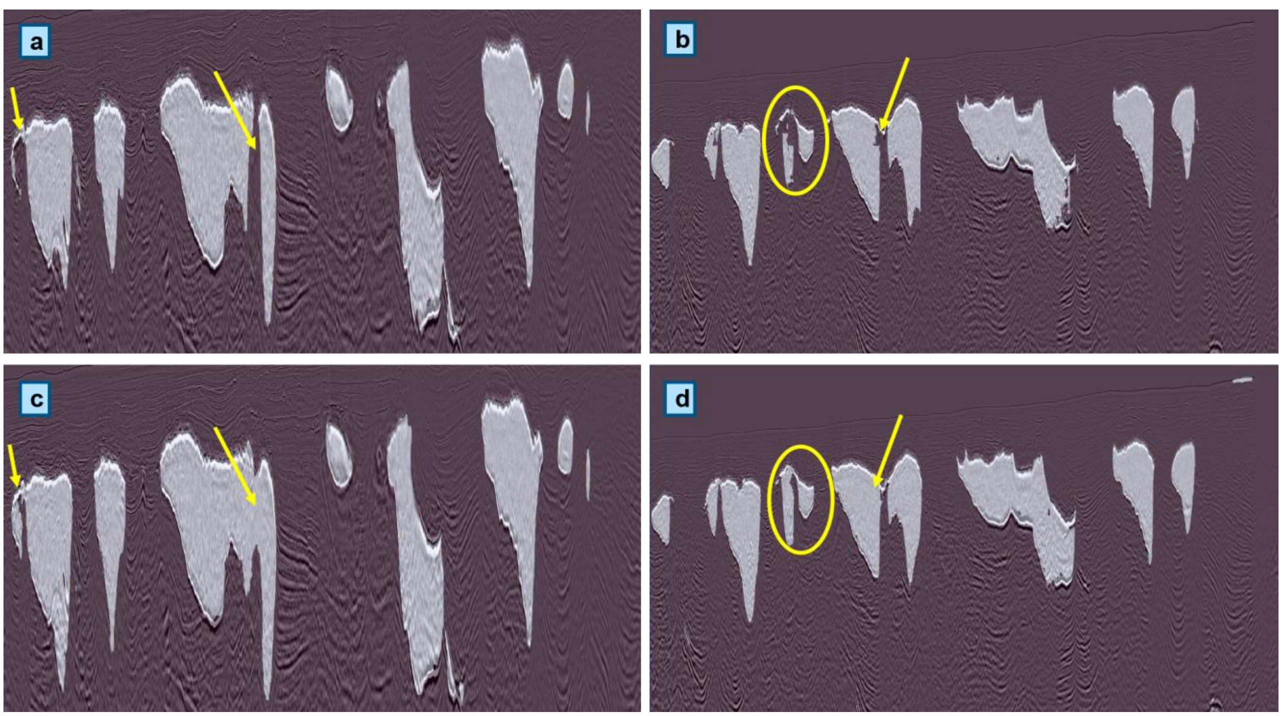


Figure 3: Salt body picking on salt flood migrated images for automated salt building using networks (a,b) without and (c,d) with the VAE branch. The picked salt masks are overlaid on the seismic.

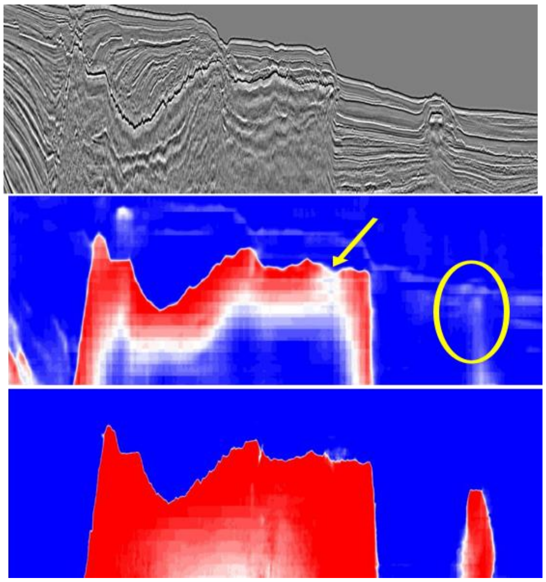


Figure 4: Salt-flood raw probabilities map picked on (a) sediment flood image by a (b) shallower and (c) deeper CNN. Shallower network has consistently more false positives and misses more subtle features easily.

the target pixel. In SaltSeg these vectors are passed to the final sigmoid classification layer and act as attention weights where the relevant portions of the segmentation maps are weighted appropriately using these vectors before the sigmoid classifier is applied.

Results

Figure 2c and 2d show the excellent reconstruction capability of the VAE branch of the network when working with complex noisy seismic images indicating that SaltSeg has learnt the identity mapping successfully. Consequently interpretation accuracy is greatly improved as is seen in Figures 3c and 3d for salt body picking on salt flood migrated image. We achieve human level interpretation accuracy and thus SaltSeg can be used as one of the key building blocks for a fully automated deep learning based salt model building pipeline. The VAE branch's reconstruction output also serves as an easy QC for the end-user to monitor the performance of the CNN. At test time if the reconstruction produced by the network shows significant differences when compared to the input image, the end-user can be certain that the CNN's training has not been optimal and suitable measures can be then taken to address the problem. The VAE branch can also be easily added to any existing encoder-decoder network and is extremely flexible to implement and use. Although in this paper we have shown results for salt interpretation only, it is trivial to extend SaltSeg to other geologic interpretation tasks like fault picking and seismic facies classification.

Conclusion

We have described a CNN architecture designed to work specifically for noisy, band limited seismic images. The CNN architecture described in this paper uses design components that can handle the specific nuances that seismic images exhibit, typically not seen in natural images. We have demonstrated that we achieve human level interpretation accuracy for salt picking during the model building stage using SaltSeg on real world seismic images. Although we have used SaltSeg primarily for salt interpretation, it is flexible enough to be easily extended for other geologic interpretation tasks.

Acknowledgements

We thank TGS management for permission to publish the results shown in this paper.

REFERENCES

Chaurasia, A., and E. Culurciello, 2017, LinkNet: Exploiting encoder representations for efficient semantic segmentation: IEEE Conference on Visual Communications and Image Processing, 1–4.

Hariharan, B., P. Arbelaez, R. Girshick, and J. Malik, 2015, Hypercolumns for object segmentation and fine-grained localization: IEEE Conference on Computer Vision and Pattern Recognition, 447–456.

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Higgins, I., L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, 2017, β-VAE: learning based visual concepts with a constrained variational framework: International Conference on Learning Representation, 1278–1287.

Milletari, F., N. Navab, and S. Ahmadi, 2016, V-Net: Fully convolutional neural networks for volumetric medical image segmentation: 4th International Conference on 3D Vision, 565–571.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-Net: Convolutional networks for Biomedical image segmentation: Medical Image Computing and Computer Assisted Intervention, **9351**, 234–241.

10.1190/segam2019-3216875.1