



# TGS SALT IDENTIFICATION CHALLENGE

*Data Science Capstone*

*Jeremy Lawson, Data Scientist*

DECEMBER 2018

# BACKGROUND

---

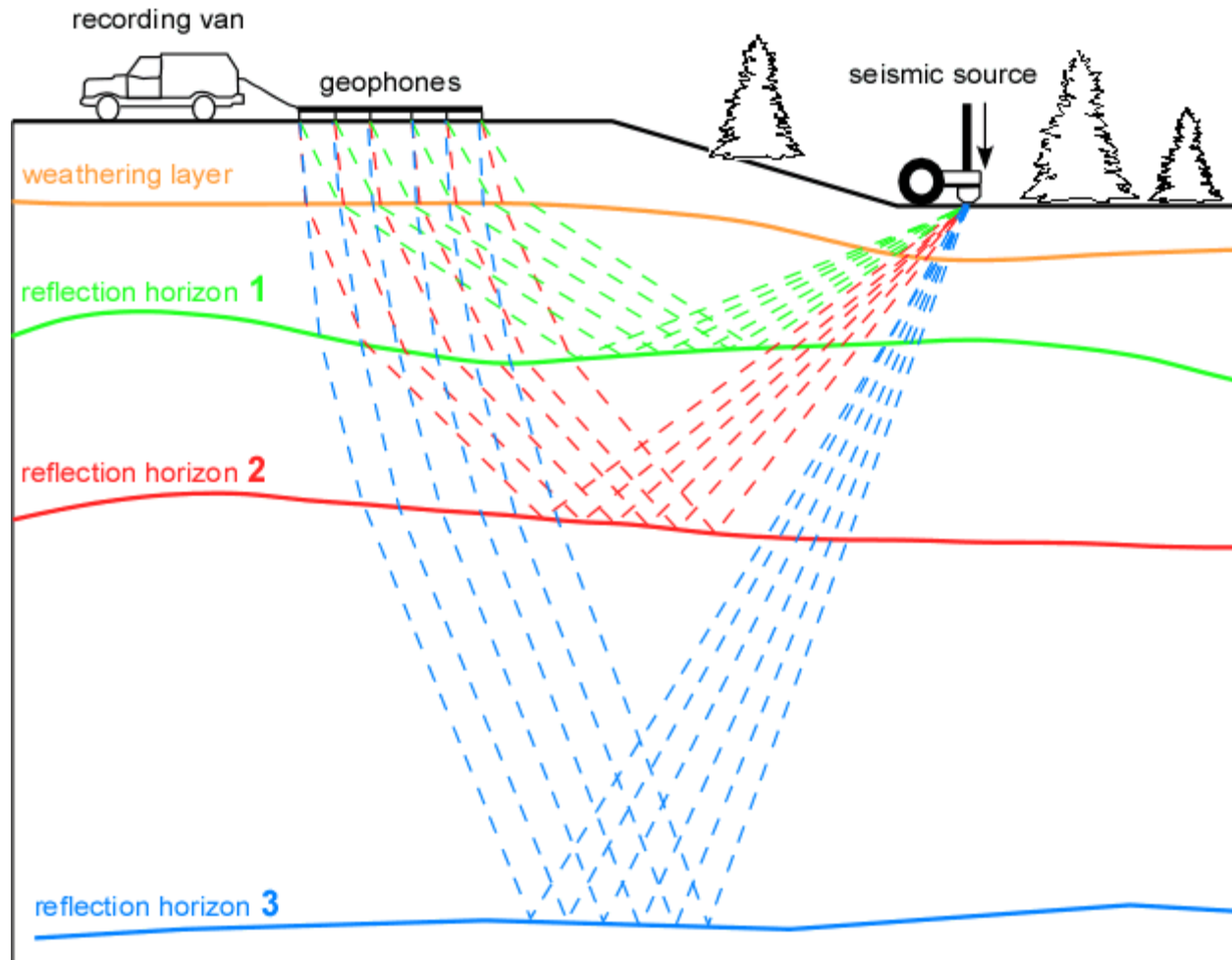
Several areas of Earth with large accumulations of oil and gas *also* have huge deposits of salt below the surface.

But unfortunately, knowing where large salt deposits are precisely is very difficult. Professional seismic imaging still requires expert human interpretation of salt bodies.

This leads to very subjective, highly variable renderings. More alarmingly, it leads to potentially dangerous situations for oil and gas company drillers



# REFLECTION SEISMOLOGY



# EVALUATION CRITERIA FOR COMPETITION

---

1. This competition is evaluated on the mean average precision at different intersection over union (IoU) thresholds.

$$IoU(A, B) = \frac{A \cap B}{A \cup B}.$$

2. The metric sweeps over a range of IoU thresholds, at each point calculating an average precision value. The threshold values range from 0.5 to 0.95, with a step size of 0.05.
3. In other words, at a threshold of 0.5, a predicted object is considered a "hit" if its intersection over union with a ground truth object is greater than 0.5

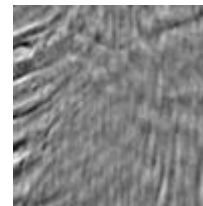
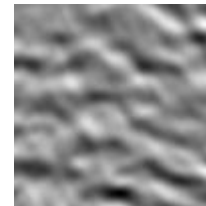
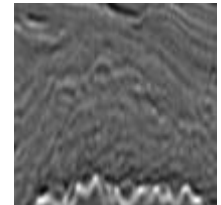
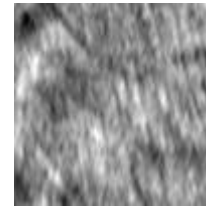
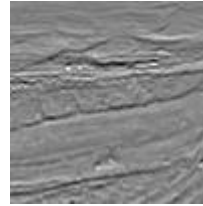
# DATA UNDERSTANDING

---

The data is a set of images chosen at various locations, chosen at random in the subsurface.

The images are 101 x 101 pixels and each pixel within each image is classified as either salt or sediment. In addition to the seismic images, the depth of the imaged location is provided for each image in a csv file.

The goal of the competition is to segment regions that contain salt.



# MODELING

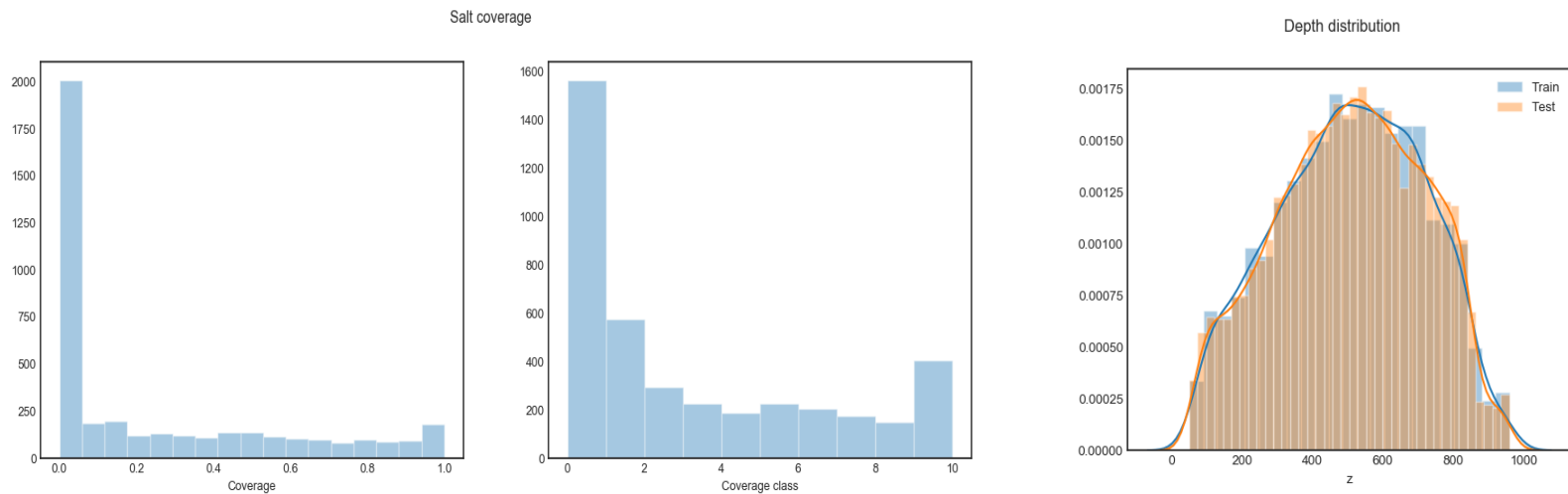
---

- Language: Python
- Framework: Keras
- Encoder: Resnet34 encoder
- Input Image size: 101x101
- 10 epochs for each stage
- Batch size = 32, min\_LR = .0001
- An 80-20 split of the train data is used for training and validation of the model.
- To increase the size of the training dataset, the training data is augmented by including images (and masks) that have been flipped along the x-axis

# MODELING

---

Figure 1 and 2 show description of the Salt Coverage and the Depth Distribution from the datasets provided.



# EVALUATION RESULTS

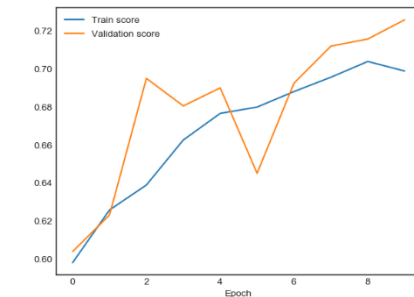
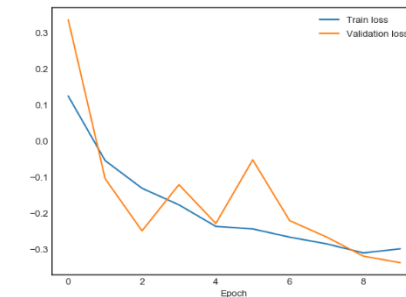
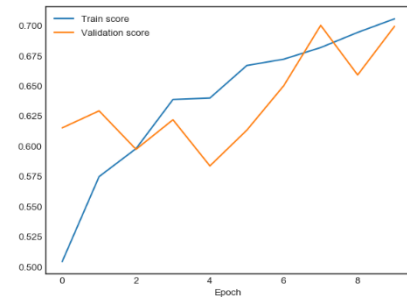
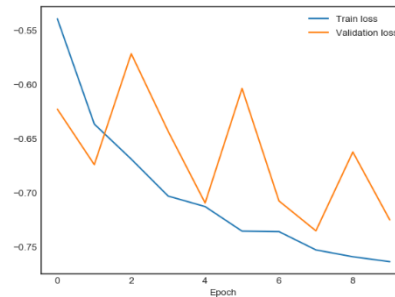
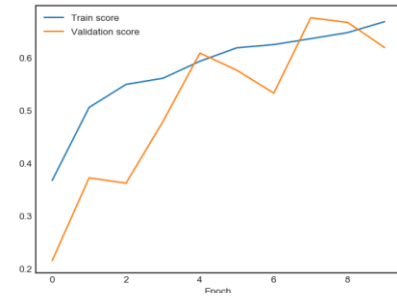
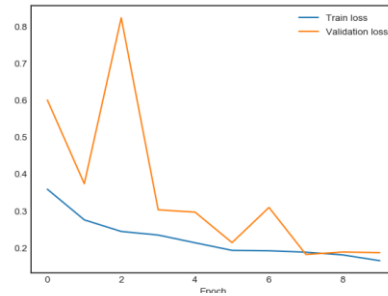
Training Overview:

Optimizer:

- The Adam optimizer

Loss:

- Binary-cross entropy
- Binary-cross entropy plus dice loss
- Binary-cross entropy plus dice loss plus mean squared error.

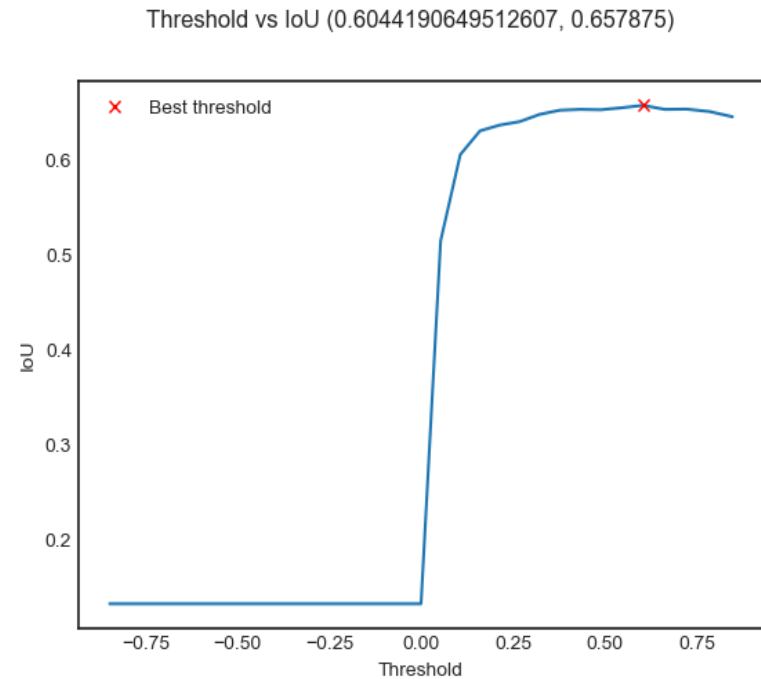




# EVALUATION RESULTS

---

Intersection over Union(IoU)  
threshold values in the range 0.5  
to 0.95



# EVALUATION RESULTS

---

Total params:	5,119,857
Trainable params:	5,112,497
Non-trainable params:	7,360

## Stage 1 Ending Results

Epoch 8/10	val_my_iou metric	0.68
Epoch 9/10	Val_my_iou metric	0.67
Epoch 10/10	Val_my_iou metric	0.62

## Stage 3 Ending Results

Epoch 8/10	val_my_iou metric	0.71
Epoch 9/10	Val_my_iou metric	0.72
Epoch 10/10	Val_my_iou metric	0.73

# EVALUATION RESULTS

---

Submission Results from Kaggle Competition

Model	Private Score	Public Score
U-net Model	0.685521	0.647333

The IoU metric improved with each stage in the code however it did not get above 70% when ran against the full test dataset for the competition.

# LIMITATION

---

Total run time =  
21.081168919139436  
hours for one iteration

Determining what is a  
signal



# THANK YOU!

---

Please Contact Jeremy Lawson at  
[Lawson.JeremyR@gmail.com](mailto:Lawson.JeremyR@gmail.com)

