**FINAL YEAR PROJECT PROPOSAL**

**TITLE**

Segmentation of Salt Bodies in Seismic Images using Semi-Supervised Convolution Neural Network.

**LITERATURE REVIEW**

Sanakoyeu *et al:* Deep unsupervised learning of visual similarities. Pattern Recognition (2018).

Favaro *et al*: Unsupervised learning of visual representations by solving jigsaw puzzles. In: European Conference on Computer Vision. pp. 69{84. Springer (2016)

Lee *et al*: Deep neural network self-training based on unsupervised learning and dropout. International Journal of Fuzzy Logic and Intelligent Systems (2017)

LeCun *et al*: Backpropagation applied to handwritten zip code recognition. Neural computation (2001)

**PROBLEM STATEMENT**

localization and delineation of subsurface salt bodies in seismic imaging has always been a challenge. The precise location of salt deposits helps to identify reservoirs of hydrocarbons, such as crude oil or natural gas, which are trapped by overlying rock-salt formations due to the exceedingly small permeability of the latter.

Modern seismic imaging techniques result in large amounts of unlabelled data which have to be interpreted. Unfortunately, the exact identification of large salt deposits is notoriously difficult and often requires manual interpretation of seismic images by the domain experts. Despite being highly time-consuming and expensive, manual interpretation induces a subjective human bias, which can lead to potentially dangerous situations for oil and gas company drillers.

The advent of convolutional neural networks (CNNs) brought significant advancements in different problems and several attempts have been made to apply CNNs in the field of seismic imaging. CNNs overcome the need for manual feature design and show superior performance on the tasks of the salt body delineation compared to the methods based on the handcrafted features.

**AIM**

The aim of this project is to build a Neural Network architecture which would be tailored for the task of salt body delineation (segment regions that contain salt and then evaluate the approach on a real-world salt body delineation seismic dataset.

**OBJECTIVES**

1. Propose an iterative self-training approach for semantic segmentation which benefits from unlabelled data.
2. Build a sophisticated network architecture which is tailored for the task of salt body delineation.
3. Evaluate our approach on a real-world salt body delineation dataset.

**DATA**

The data for this competition represents 2D image slices of 3D view of earth’s interior gotten from Kaggle Dataset (TGS Salt Identification Challenge dataset) It was collected using reflection seismology method (similar to X-ray, sonar, and echolocation) For this reason, input data is a set of single-channel grayscale images showing the boundaries between different rock types at various locations chosen at random in the subsurface large-size images were transformed into 101 *×* 101 pixel crops by the organizers. Further, each pixel is classified as either salt or sediment and binary masks are provided.

The whole dataset has been split into three parts: train, public test, and private test. The train set consists of 4000 images together with binary masks and is used for models developing. The public test set has around 6000 images and is used for evaluating the models during the competition. Lastly, private test set has around 12000 images and is used to determine the final competition standings. Overall, the test dataset contains 18000 unlabelled images (public + private test) which we can use for self-training.

**METHODOLGY**

By employing an ensemble of two U-Nets with ImageNet-pretrained encoder backbones: U-ResNet34 and U-ResNeXt50. The output of the ensemble is the average of the predictions of two models in the ensemble. All images will be resized to the size of 202 *×* 202 pixels and then padded to the size of 256*×*256 pixels

Additionally, to get a more robust ensemble at the end of every round I will average the predictions of 4 snapshots, which will be saved every 50 epochs. This strategy will yield better results than using only confident pseudo-labels.

During the first self-training round, I will train the ensemble on the provided 4000 labelled images and generate 18000 pseudo-labels for unlabelled images. At rounds 2 and 3, I will train the network for *T* epochs solely on the pseudo-labelled data and then fine-tune for another *T* epochs on the ground truth labelled training images.

After each stage, I will obtain 4 network snapshots for each of 5 folds giving 20 snapshots in total for a single network architecture. I will also use an ensemble of U-ResNet34 and U-ResNeXt50, which will result in 40 models in total and will be combined together for inference using the average voting.

For the final prediction on the test set, I will use an ensemble of Round 2 and Round 3 models, which will give the best performing results on the test and real datasets.

**RESULT**

Afterapplying a sophisticated network architecture which is tailored for the task of salt body delineation, and then applying on a real-world salt body delineation dataset. The result should be:

* A mapping of Seismic Images relating the test and Validation set perfectly correlated.
* A display of seismic Images showing Salt and non-salt Images.
* An accuracy score with satisfactory and convincing evidences of a near to perfect work.
* Application of this architecture should also work on any Salt data from any Seismic Images.

**CONTRIBUTION TO KNOWWLEDGE**

The exact identification of large salt deposits is notoriously difficult and often requires manual interpretation of seismic images by the domain experts. Despite being highly time-consuming and expensive, manual interpretation induces a subjective human bias, which most often lead to potentially dangerous situations for oil and gas company drillers.

The advent of convolutional neural networks (CNNs) brought significant advancements in different problems and several attempts have been made to apply CNNs in the field of seismic imaging. CNNs overcome the need for manual feature design and show superior performance on the tasks of the salt body delineation compared to the methods based on the handcrafted features.

**DURATION OF PROJECT**

The Time duration I proposed for this project is Two Months.

**FINANCIAL IMPLICATION**

* Cost of getting the Data involves registration on the Kaggle site, and upgrading to certain level which include a payment of $50, around 35,000 in naira.
* Cost of Extraction and Collation involves some APIs like TensorFlow, Pytorch, Keras, and XGBoost which requires upgrading to certain level with payment of $60, 45, 000 in naira.