Machine Learning Engineer- Oil Seep Detection Exercise

This is an image segmentation exercise to test your ability to construct and evaluate machine learning models.

The objective of this exercise is to produce a deep convolutional neural network (DCNN) model and an evaluation metric. We are not expecting you to produce a state-of-the-art machine learning model.

If your chosen model performs poorly by your selected metric, do not worry - this is not what we are testing. Which model you use, and how you evaluate, is up to you.

The choice of model is not important (although we will assume that when you choose a model, you understand what it is and how it works). Your solution should be simple, but sensible: you should be able to explain why it tests something of impact to the problem.

Data

The data is a set of synthetic aperture radar (SAR) images chosen at various locations. The images are 256 x 256 pixels and each pixel is classified as non-seep (0) or 7 classes of seeps (1-7). The SAR images and their corresponding masks are saved as .tif files with the same names, but in separate folders, train_images_256/ and train_masks_256/. The objective of the exercise is to segment regions that contain seeps, and as an optional task to classify the seeps.

Tips and Clarifications

▪ We are not looking for a model that performs well: we are looking to see that you can build a sensible model with a sensible evaluation.

▪ You should aim to spend about 4 hours on the problem, and submit a solution with no more than a couple of hundred lines of code.

▪ If you are struggling to make something work with the volume of data present, you can subsample or resize the images. If you are struggling to implement something that deals with this volume of data, do you know of a way to deal with it in theory?

▪ You can use any programming language you like to solve the problem: pick a language suited to the task, and one you are comfortable with.

▪ You are strongly encouraged to make use of third-party libraries for model building and evaluation, rather than writing your own, unless you specifically need to do something with no library support.

▪ We will reimburse you up to a maximum of $68.00 for using GPU(s) from cloud providers (AWS, Google cloud and Azure).

After you have completed the seeps exercise and received feedback please email your receipt/order confirmation to Lauren.Metz-contractor@CGG.com along with your bank details.