Imaged Reality

Company Background

We built our platform for a purpose: to enable geologists and engineers around the world to study locations that they couldn't visit in person while connecting expertise and data in an immersive environment. Imaged Reality comprises an internationally diverse team of experts in geology, software development and gaming technologies.

Today this is a reality. We are partnering with experts to bring their knowledge wherever it is needed, with the help of our immersive technology that brings the landscape to life. Whether you're working in industry, training or academia, we can help you to leverage the power of virtual reality to work in a safer, smarter and more inclusive, innovative way. See our website here.

Project Description

Problem

Drill core images are commonly used by geologists to make inferences about subsurface features, and to guide decision making in further exploration, drilling and extraction scenarios. Imaged Reality is leading the industry with tools to visualize core image datasets - in order to make fast, accurate analyses, and to facilitate knowledge sharing. However, this process often relies on judgements made by the analyst, and is subject to bias & error, depending on experience level.

Solution

Machine Learning will be used to train an inference model that predicts the type of lithology at regular intervals when applied to drill core (slab type) images. Training data will be collected and annotated with associated lithology information (ie rock type; *Biotite, Dolomite, Iron Ore,* etc. and corresponding depth values). The Team will then develop, train and iterate until an accurate prediction model is obtained. A *transfer-learning* approach may be used to increase the efficiency of the training process. Similar projects in academia have achieved high levels of accuracy when using this technique. A recent example is available at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270826

Prototype Expectations

The completed project should be modular, and follow existing best practices for Machine Learning development. This includes;

- use of source control software (eg Git),
- hyper-parametrization of key modifiers,
- well-organized training datasets,
- use of appropriate compute resources for training,
- logs of all training runs, along with checkpoints

- portability & extensibility of the final model

Technologies

Several frameworks exist for Machine Learning development, and we would prefer the Team to use either *Pytorch* or *TensorFlow*. Additional data science libraries such as *Keras, Scikit-learn, Numpy* and *OpenCV* may also be necessary - in order to accelerate development and provide extensibility for the codebase. However, the specific choice of which (if any) of these additional libraries are to be used - will be left up to the Team - depending on their experience and preferences.

Collection and annotation of training datasets will require the use of a Web Scraping framework such as *Beautiful Soup* or *Selenium*. Since this is not a geology problem per-se, assistance will be provided by Imaged Reality staff for this aspect of the project, in order to allow the Team to focus on the core problem.

Imaged Reality will provide all compute-intensive resources for the Team to use during development. Depending on experience and preference, this will be made available as either Jupyter Notebook-linked GPU instances in the cloud (AWS Sagemaker), or via a dedicated CLI.

Student Requirements

The Team should have some familiarity with the basic principles and practice of Machine Learning development. Experience with either of the aforementioned frameworks (Pytorch or TensorFlow) is strongly recommended, as is a familiarity with the iPython / Jupyter Notebook system. Specific experience with Computer Vision, Production ML pipelines and Web Scraping would be nice-to-have, but is not a requirement.