Open-source Python Stack and Tools for Geoscientific Image Analysis and Interpretation -From research to deployment

Mustafa Al Ibrahim[1]

[1]Saudi Aramco

Abstract

Geoscientific images tend to be multi-dimensional in space, time, and data type. Increasingly, they are large in terms of size and amount. Integration of this data requires developing specialized algorithms and workflows. These developed workflows need to be accessible to geoscientists and engineers for their daily use. This typically means the implementation and transformation of a research-oriented study, typically developed in codes and digital notebooks, to: 1) graphical user interface, and 2) guided workflows. We present how open source scientific libraries can be utilized to accelerate research, development, and deployment of geoscientific image processing technologies in operational settings. The proposed Python-based stack is composed of layers that rely on each other: 1) base libraries such as vector manipulation libraries numpy and dask, 2) image processing libraries such as scikit-image and open-cv, 3) machine learning libraries pytorch and scikit-learn, 4) visualization library/tool napari, and 5) custom developed modules/plugins. Solutions are packaged using briefcase to produce an easy to use installer for users. An important aspect of this stack is the use of lazy-loading implemented using dask. This allows for processing and visualization of large files. Outputs are saved in open-source lazy-loading friendly formats such as zarr and ome-zarr. Since the development stack uses the same libraries as research, it is straightforward to transfer solutions from the research environment, e.g., notebooks, to operations, i.e., graphical user interfaces. We show that using this stack, a number of technologies can be developed. Specifically, it is straightforward to develop modules for loading, visualizing and processing of seismic, 2D thin sections, 3D CT-scans, and core and cutting images. The extensibility of the used open-source visualizer, napari, allows for first-party grade modules. As a detailed application example, we show that different types and sizes of thin section rock images can be easily visualized. Specialized modules for thermal maturity estimation from palynomorphs using palynomorphs darkness index, and thermal alteration index use image processing and machine learning as backend to automate the estimation partially or completely. Semantic segmentation modules use convolutional neural networks to extract rock components. Overall, the use of this stack accelerated the implementation of advanced research in daily operational workflows.