Classification of Jurassic palynomorphs using an expert-trained convolutional neural network

De Bakker, M.P.[1], Houben, A.J.P.[2], Abdul Fattah, R.[2], Verreussel, R.M.C.H.[2] and Williams, R.[3]

Palynomorphs of terrestrial and marine origin are critical elements in paleo-environmental and biostratigraphic studies, both in an academic and industry context. Palynological data acquisition is a very time-consuming process, as experts need to taxonomically classify and subsequently quantify a palynological association. This not only limits the statistical significance (n=limited) but potentially also introduces analyst-based bias.
Over the past few years a technical revolution in (I) photographic digitization of microscope slides and (II) the open-source development of image classification and localization algorithms, coupled with ever increasing computational capacity, now allows for possibilities to automate palynological classification and quantitative analysis.

Here we present the results of a pilot study aimed at the automated classification of palynomorphs in a rich and diverse palynological association from the Kimmeridgian (Late Jurassic) from the Norwegian Sea (Well 16/3-2). This slide was first scanned at visual high resolution at using a 40x objective at the Norwegian Petroleum Directorate using a Pannoramic 1000 pathological scanner. Scanning was performed at five focus-levels. Subsequently, the eight focus levels were flattened using the 3D-HiStech software package, resulting in a single image of the slide, with a bigger depth of field, allowing entire palynomorphs to sharp in focus, independent of their 3D-geometry.

An expert in Mesozoic palynology labelled the palynomorphs in the slide into 24 different taxonomic groups. These include 21 genera of organic-walled dinoflagellate cysts. In addition, miospores, bisaccate pollen grains and acritarchs were labelled as distinct palynological categories. For each class, at least seventy specimens were labelled. In total the training set consists of 1250 labelled palynomorphs.

A Single Shot Multibox Detector (Liu et al. 2016) algorithm was trained for classification and localization of polynomorphs. A SSD is a supervised machine learning algorithm based on convolutional neural networks. Once trained, this algorithm provides bounding boxes and labels for found objects in the image.

Because of the size of the palynological slices and the high number of parameters to be optimized in the model, it was not feasible to train the network on the full slices. Therefore, we randomly cropped the original slice into smaller images. We did not use any data augmentation techniques in this pilot.

The dataset was split into training, testing and validation sets. A pre-trained ResNet model was used as a base model, retrained using the polynomorph images. The overall classification accuracy is 0.63 on the testset. However there is high variability between the different classes.

We think this pilot study shows a potential for the automated classification and localization of polynomorphs. The algorithm already shows a decent performance on a limited dataset, without using any hyperparameter optimization or data augmentation procedures. It is to be expected that with such relatively easy additional procedures, the performance can be further improved.

[1] TNO Monitoring and Control Department, The Hague, The Netherlands
[2] TNO Geomodelling Department, Utrecht, The Netherlands
[3] Norwegian Petroleum Directorate (NPD), Stavanger, Norway