Contribution ID: **14** Type: **not specified**

Pl@ntNet: towards the recognition of the world's flora

Wednesday, 10 October 2018 08:45 (30 minutes)

Automated identification of plants and animals have improved considerably in the last few years, in particular thanks to the recent advances in deep learning. In 2017, a challenge on 10,000 plant species (PlantCLEF) resulted in impressive performances with accuracy values reaching 90%. One of the most popular plant identification application, Pl@ntNet, nowadays works on 18K plant species. It accounts for million of users all over the world and already has a strong societal impact in several domains including education, landscape management and agriculture. The big challenge, now, is to train such systems at the scale of the world's biodiversity. Therefore, we built a training set of about 12M images illustrating 275K species. Training a convolutional neural network on such a large dataset can take up to several months on a single node equipped with four recent GPUs. Moreover, to select the best performing architecture and optimize the hyper-parameters, it is often necessary to train several of such networks. Overall, this becomes a highly intensive computational task that has to be distributed on large HPC infrastructures. In order to address this problem, we used the deep learning framework Intel CAFFE coupled with Intel MLSL library. This experiment was carried out on two french national supercomputers, their access was offered by GENCI. The first experiment was carried out on Occigen@CINES, a 3.5 Pflop/s Tier-1 cluster based on Broadwell-14cores@2.6Ghz nodes. The second uses the Tier-0 «Joliot-Curie»@TGCC, a BULL-Sequana-X1000 cluster integrating 1656 nodes Intel Skylake8168-24cores@2.7GHz. We will report our experience using these two platforms.

Presenter: JOLY, Alexis (Pl@ntNet)

Session Classification: Plenary