



Contribution ID: 78

Type: **not specified**

Machine Learning at Diamond Light Source: Past, Present and Future

Wednesday, 13 November 2019 16:25 (25 minutes)

Around 8 years ago, the B24 beamline at Diamond Light Source identified an issue with its potential data collections, that we didn't have an immediate answer for. The issue was that although the planned beamline could collect a Cryo Soft X-ray Tomography image of sub cellular structure with amazing resolution in a matter of minutes, due to the complexities of radiation damage and acquisition geometry, the data was noisy with several significant artefacts. These issues manifested themselves when trying to derive quantitative information from the datasets, which generally requires the process of semantic segmentation. Unfortunately, at the time, this process could take an expert researcher weeks' to achieve, so without a significant number of PhD and postdocs manually annotating datasets, the full potential of the beamline may not be achieved. Given this complex problem, Diamond looked to the relevant UK academic communities for help. Within a couple of years a joint PhD student with Nottingham Universities Computer Vision Lab, who had experience with segmentation of complex biological volumes of root structures, was appointed. They immediately researched the problem, and very quickly identified that standard algorithmic and modelling approaches were not going to work, but that a Machine Learning approach may offer the best chance of success.

Now the resulting software is in widespread use around Diamond, we have new PhD students looking into Deep Learning for image segmentation, and we are starting a 3-year stretch of funding to further expand the approach using the Citizen Science Platform Zooniverse. In complex biological segmentation, the future seems bright in the area of Machine Learning, and although we have not yet managed to remove the human in the loop, we have achieved a five-fold reduction in the amount of time they have to spend manually working on the problem. These successes, along with an expanding presence on campus and globally through the general media and publications, have also drawn a much larger community at Diamond. These researchers are looking to harness these methodologies in many diverse ways, to analyse, predict and accelerate the work they do on a day to day basis, for example the new DIAD (Dual Imaging and Diffraction) beamline is investing heavily in ML to help guide users to areas of interest in samples which are changing over time. In addition, we are starting to see third party tools that make use of Machine Learning, becoming easily available and simple to integrate, allowing beamlines to easily tackle complex issues such as crystal finding in Macro molecular Crystallography and particle picking in cryo Electron Microscopy Single Particle Analysis.

Diamond Light Source currently collects around 5 Peta-bytes of Data every year, a value which is increasing at an alarming rate. To continue to publish the work conducted at such a facility, involves distilling potentially tens of Tera-bytes of data down to the refined information that can be represented in a tens of Mega-byte journal publication. We already use Machine Learning to help our researchers with this challenging goal, and can only see its use increasing in the future, as the methodologies, applications, and ingenuity of their application increases.

Presenter: Dr BASHAM, Mark (DLS)

Session Classification: Afternoon 2