Abstract: Automated classification of In Situ Ichthyoplankton Imaging System (ISIIS) images using Convolutional Neural Nets on parallel computing infrastructure, Y2

The In Situ Ichthyoplankton Imaging System (ISIIS) is a towed underwater imaging system capable of detecting individual organisms and particles ranging from a few hundred micrometers up to 13 cm. The system has been deployed in near-transparent ocean waters where the image frames (13x13x50 cm) typically yield a few hundred million individual segments of planktonic organisms and detritus particles over the timeline of a project. An image analysis pipeline that combines image flat fielding, Region of Interest (ROI) detection and segmentation, as well as a Convolutional Neural Net classifier has been tested and fully implemented in previous projects with excellent results. However, with the interest of understanding the ecology of the highly productive northern California Current Ecosystem, we are facing the challenge of classifying an unprecedented number of images (a two-order of magnitude higher number of images than any previous imagery collection). Doing so with the GPU resources permanently available to us locally (2 PCI-based k80s) would require years of GPU time. This is likely to become a continuous challenge when studying highly productive waters with high ecological and economic importance. We are collaborating with an international team of plankton ecologists and computer vision experts to create a standard image-processing pipeline (Belmont Forum (NSF) funded project: World Wide Web of Plankton Image Curation), and we were recently funded by the NOPP (NASA) to work on a Marine Biodiversity Observing Network. Our Xsede imaging pipeline is now operational, segmenting our images locally and then porting them to Xsede for classification by our sparse Convolutional Neural Network. Classification files are then sent back to our local infrastructure for further QA/AC. This is a major development since the last request which is detailed in the progress report document. While these results are very good, and we classified multiple hard drives worth of data, we would like to request more time and resources to classify all current data, as well as data from two additional cruises that will be carried out this year. Hence, we propose to continue advancing the use of Xsede's available computational resources to expedite the automated classification of plankton imagery in order to generate meaningful ecological data, and to address important questions regarding plankton production and distributions in the California Current.