Building platform for medical image segmentation using High Performance Computing

Petra Svobodova, Khyati Sethia and Petr Strakos

IT4Innovations, VSB – Technical University of Ostrava, Ostrava, Czech Republic

Abstract:

Medical image processing enables doctors to see the anatomy of the human body and discover disorders. Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) are frequently employed as data sources. Machine learning, specifically deep learning (DL), is used as the current approach to medical image processing and analysis. The goal is to shorten the time it takes to acquire a reliable medical diagnosis and, as a result, enhance healthcare in general.

A large amount of data and computational power is needed especially for training neural network models from scratch. The idea of our solution is (i) to provide an AI-based annotation service that uses the state-of-the-art DL algorithms for automatic segmentation of desired tissues and (ii) collect the data after automatic segmentation and validation by doctors and provide High Performance Computing (HPC) based training of new models or fine-tune the existing models. Resulting models are then used again in the step (i). The idea of such medical service is highly demanded. Using our approach, personalized medicine can be achieved in fast image to result process. Proposed solution provides ways for automatic segmentation of abdominal organs, bones, and even hepatic arteries that can be viewed as highly detailed 3D models.

The entire concept is an outcome of the collaboration between a local hospital and our HPC center. The local hospital runs frontend software that mediates the interaction between the doctor and the data. It enables doctors to upload the data, view it in its raw form, and perform inference by accessing the computational resources of a cluster that processes the images and returns the results. An open-source 3D Slicer is used as the frontend viewer, and it creates a portable and lightweight solution since all the computation is performed remotely at the backend that runs on the HPC cluster. We developed the required extension for the viewer to allow the communication with the backend. At the backend we use the NVIDIA Clara Train SDK. It contains Al-Assisted Annotation (AIAA) and Training Framework APIs. Clara framework uses its tools in conjunction with open-source Medical Open Network for AI (MONAI) component that is specifically denoted to DL in healthcare imaging.

We use Singularity to provide Clara as a containerized solution that can run effectively on HPC hardware. In Clara we started with the pre-trained models from NVIDIA. Such models are stored in the so-called Medical Model Archive (MMAR) which establishes a generic format to manage configuration files and the data. We have extended the model's capabilities to better match the needs of doctors by writing specific data processing scripts. This is possible by using bring your own components (BYOC) concept of Clara. The AIAA part of Clara provides annotation server and offers methods for labelling datasets such as segmentation, annotation, and deep grow. The AIAA can automatically segment specific tissues based on the pre-loaded models on the server and then share the results with the frontend client. The refinement of the generated masks can be done by the doctors directly in the frontend. The resulting high-quality annotated data can be sent to the cluster and used to provide robust DL models. The DL models are trained by utilizing the cluster resources in the form of multi-GPU compute nodes. Trained models are provided to the frontend to perform model inference on a single GPU based visualization server on the cluster. Only anonymized image data is transmitted between the hospital and the HPC centre. Besides that, all connections are encrypted by secure shell (SSH).