Bardia Mojra – 1000766739

02/28/2022 – Assignment 02 – Application 02

Supercomputer #4 – JUWELS Booster Module Supercomputer

Super-Resolution of Large Volumes of Sentinel-2 Images

with High-Performance Distributed Deep Learning

In this paper, the authors use the JUWELS booster module high-performance computing (HPC) system to train a deep learning (DL) model for satellite super-resolution (SR) applications. More precisely, they propose a novel distributed DL model for remote sensing (RS) image SR application of data gathered by the Sentinel-2 satellite. They use GPU accelerators and synchronous data parallelism to achieve state-of-the-art performance while achieving substantial speed-up during model training. For their model, they mostly rely on the self-attention mechanism, residual learning while keeping the model size relatively small.

The spatial resolution of the satellite images is measured by Ground Sampling Distance (GSD) which represents the distance between two sequential pixel centers measured on the ground. High-resolution satellite imagery and RS applications have become an integral part of life in the 21st century and the demand for higher resolution steadily increases. The problem arises from hardware limitations in various ways and the inability to make frequent updates. Sensors and communication hardware onboard satellites are examples of real-life bottlenecks that limit high-resolution image capture and transmission. Thus, the need for super-resolution post-process applications is also increasing. Moreover, satellites acquire vast amounts of data that are hardly supported on HPC platforms. For example, the twin Sentinel-2 orbiters capture up to about 23TB/day of multispectral images. It could take weeks or months to train and optimize a DL model with a month’s worth of data if it were to run even on the most powerful desktop computers. Thus, an HPC-based solution becomes even more necessary as the significant training speedup allows researchers to explore more innovative models with different structures in pursuit of an optimal DL model. In this paper, the authors propose an HPC-based DL model for satellite imagery super-resolution applications which is smaller in relative size and state-of-the-art performance.

Sentinal-2 is deployed by the European Space Agency to capture and communicate multi-spectral optical observations over land and sea surfaces. It is considered a high revisit frequency probe which makes it capable of capturing temporal data. It has 13 spectral bands on board that can be divided into three sets based on their GSD, set A 10m bands, set B 20m bands, set C 60m bands. It is important to note set A has the highest resolution and is used as ground truth data. For the purpose of this study, they super-resolve 20m and 60m bands to 10m bands from A, the two models are designated S2X and S6X respectively.

Both models share similar network architecture with a few differences that are mentioned later in this document. The S2X network begins with upsampling the 20m bands by two times using bilinear interpolation. Then, the upsampled signal is fused with the 10m bands through a band fusion module. Then, the network proceeds with a residual self-attention module (RSA), followed by a final convolutional layer that learns the residual difference between the high-resolution 10m bands and the interpolated and upsampled 20m bands. Convolutional Neural networks cannot learn global patterns among input pixels due to the nature of their convolutional process. To mitigate this and learn global patterns, six residual blocks were deployed in the SRA module, with a self-attention module to capture long-range dependency over the input feature maps. The S6X model has similar architecture, except it upsamples by a factor of six and has an additional input channel from the 20m bands, as well as 60m bands.

Per published literature on the topic, there are two general paradigms for parallelization of training for DL models, *model parallelism* and *data parallelism*. In this paper, the authors apply *synchronized data parallelism* to speed up the training process, which specifically takes advantage of large GPU on clusters on an HPC system. They split batches of data and divide the workload among the available GPU workers to be processed in parallel. At each optimization epoch, they synchronize by averaging and aggregating the gradients over multiple workers. They use a library called, Horovod, which deploys a *Ring Reduction Mechanism* to scale up, which has been proven to be *bandwidth optimal.* To improve the model convergence of mini-batches when the training is scaled, they present a *modified linear learning rate scale rule* to avoid suffering from *explosive loss*. The HPC scale rule states the initial learning rate should be increased by the same factor as which the mini-batch size has scaled up and decayed to half every SGD iteration.

To train and evaluate the proposed models, the authors use the Dsen2 dataset and use the Wald’s protocol to generate the ground truth. The mentioned models are implemented with Tensowflow with an ADAM optimizer and a modified linear learning rate scale rule of 64,000. The models were trained on the JURON and JUWELS HPC systems. The JUWELS supercomputer (which is the subject of this assignment) is the world’s 8th best HPC system with each of its compute nodes are equipped with four Tesla V100 GPUs that are interconnected via NVLink with an all-to-all structure.

|  |
| --- |
| Table 1 |

The authors use the following 6 metrics in the evaluation of their proposed models, the root-mean-square error (RMSE), signal-to-reconstruction error (SRE), spectral angle mapper (SAM), Erreur Relative Globale Adimensionnelle (SSIM) and peak signal to noise ratio (PSNR). They compare the results with naive bicubic interpolation and the results provided the model published alongside the Dsen2 dataset. They show that they beat state-of-the-art performance when trained both models in 4 GPU configuration while training four times faster. The performance begins to suffer slightly as the GPU configuration is increased to 8 and 16 while speeding up as expected. Table 1 shows the performance S2X model with different GPU configurations as well naive bicubic interpolation and DSen2 for reference. The drop in performance is attributed to communication overhead among the GPUs. The S6X model also shows similar performance.

Thus, the authors were able to speed up the DL training process for a multi-spectral super-resolution application using the JURONS and JUWELS HPC platforms and reduce actual training time significantly. They were able to train their DL models faster and achieve state-of-the-art performance in a 4 GPU configuration. Moreover, they showed that HPC-based DL models can utilize higher GPU counts to achieve higher speedup rates for training without suffering from a significant decrease in performance.

Corresponding Authors:

[zhang@aices.rwth-aachen.de](mailto:zhang@aices.rwth-aachen.de) (R. Zhang), [g.cavallaro@fz-juelich.de](mailto:g.cavallaro@fz-juelich.de) (G. Cavallaro), and [j.jitsev@fz-juelich.de](mailto:j.jitsev@fz-juelich.de) (J. Jitsev)

References:

1. Zhang, Cavallaro, and Jitsev, “Super-Resolution of Large Volumes of Sentinel-2 Images with High Performance Distributed Deep Learning,” International Geoscience and Remote Sensing Symposium (IGARSS), 2020, PP. 617–620, IEEE.