

Machine Learning-based Data Compression

Saturday 20 July 2024 15:55 (17 minutes)

“Data deluge” refers to the situation where the sheer volume of new data generated overwhelms the capacity of institutions to manage it and researchers to use it. This is becoming a common problem in industry and big science facilities like the MAX IV laboratory and the LHC.

As a solution to this problem, a small collaboration of researchers has developed a machine learning-based data compression tool called “Baler”. Baler allows researchers to design lossy compression algorithms tailored to their data sets via an easy-to-use pip-package. This compression method yields substantial data reduction and can compress scientific data to 1% of its original size.

Baler recently performed compression and decompression of data on FPGAs, which extends Balers reach into the field of bandwidth compression. This contribution will bring an overview of the Baler software tool and results from Particle Physics, X-ray ptychography, Computational Fluid Dynamics, and Telecommunication.

Alternate track

1. Computing, AI and Data Handling

I read the instructions above

Yes

Primary authors: GALLEN, Axel (Uppsala University (SE)); DOGLIONI, Caterina (University of Manchester (GB)); EKMAN, Per Alexander (Lund University (SE)); JAWAHAR, Pratik (University of Manchester (UK - ATLAS))

Presenter: GALLEN, Axel (Uppsala University (SE))

Session Classification: Computing and Data handling

Track Classification: 14. Computing, AI and Data Handling