

Coleman Nichols and Jon Calhoun (advisor)

Holcombe Department of Electrical and Computer Engineering - Clemson University

Motivation/Goals

- Los Alamos National Labs is analyzing an additive manufacturing process
- They are taking pictures of this process to train a Machine Learning Algorithm
- The number of images needed far exceeds the data storage methods warranting data reduction methods
- The goal is to compare different reduction methods based on 3 different metrics
 - PSNR – (Peak Signal to Noise Ratio)
 - Compression Rate
 - Compression Ratio

Reduction Methods

- Hybrid Data Sampling
 - Samples a percentage of the original data with complete resolution based on “importance”
 - Forms a reconstructed version of the data from these saved segments
- SZ Lossy Compressor
 - Given error bounding method and error bound, will compress and decompress the data with slight error
- BLOSC Lossless Compressor
 - Compresses and decompresses data with no error whatsoever

Important Formulas

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

$$Compression\ Ratio = \frac{Uncompressed\ Size}{Compressed\ Size}$$

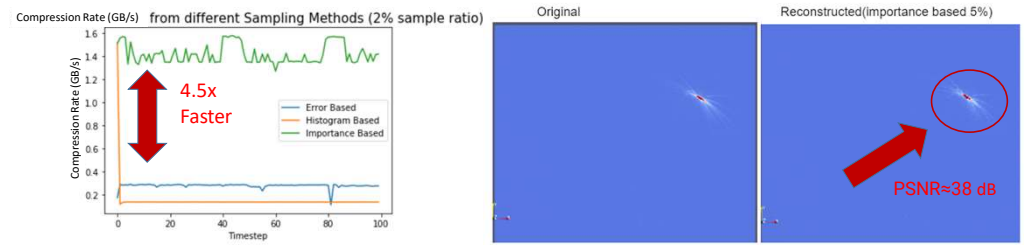
$$Compression\ Rate = \frac{Uncompressed\ bits}{seconds\ to\ compress}$$

Software Versions

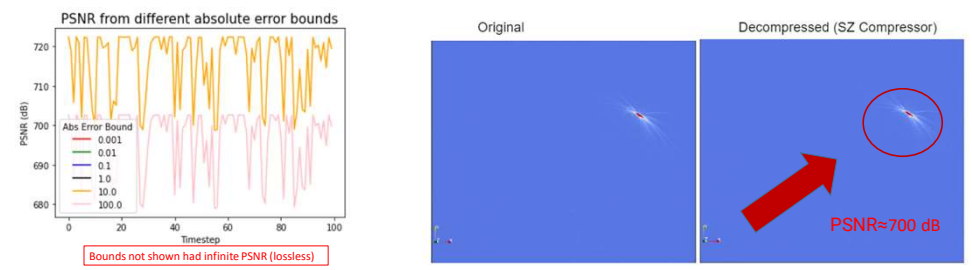
- Software Versions
- Libpressio 0.79.0
 - Python 3.9.2
 - OpenMPI 4.0.3
 - CUDA 11.5.0

Lossy Methods

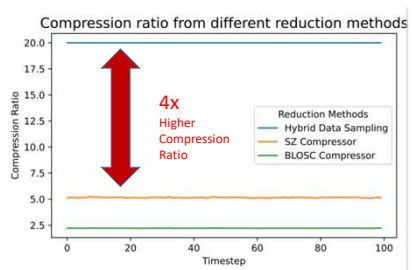
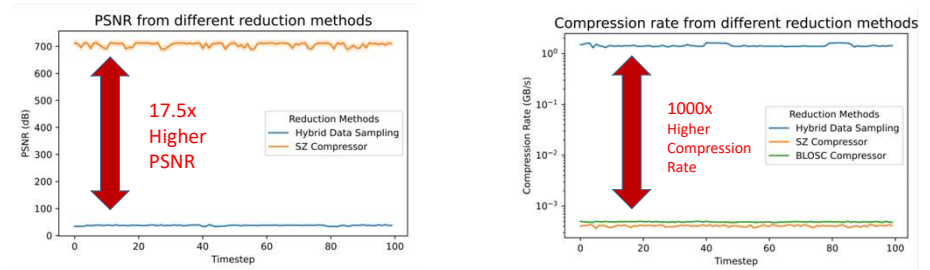
Hybrid Data Sampling



SZ Compressor

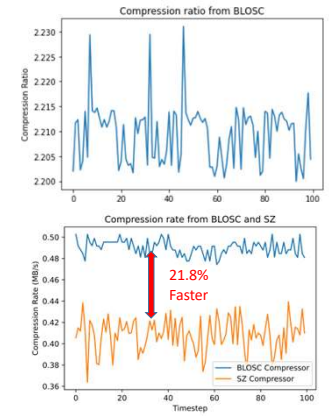


Comparing All Methods



Lossless Method

BLOSC Compressor



Future Goals

- Finding an efficient way to combine both the Hybrid Data Sampling method with the lossy/lossless compressors
- Could be specific to LANL's dataset or more general
- Utilize the compressor's accuracy with the Hybrid Data Sampling's speed

Acknowledgements

Clemson University is acknowledged for generous allotment of compute time on the Palmetto cluster. This material is based upon work supported by the National Science Foundation under Grant No. SHF-1910197 and SHF-1943114.

Dataset From Los Alamos National Labs LA-UR-21-32202

Key Takeaways

- Importance Based Hybrid Data Sampling was fastest
- Hybrid Data Sampling offered relatively low PSNR
- BLOSC compressor could cut image size in half with no error (faster than SZ)
- SZ could reduce image by 5x with no error