

Statement of motivation

By Aman Singh Thakur

Ever since I saw the first season of the TV series Silicon Valley, I have been fascinated with Data Compression. During my undergrad, I understood many loss-less compression techniques like Huffman coding, etc, but in this age of data explosion, I see loss-less Compression as the future of data storage. As part of Google Summer of Code 2019, I collaborated with CERN scientist Dr. Spencer Gessner and the CERN SWAN team to compress 12TB of data into ~150GB metadata index, allowing for smoother times-series image and data analysis [\[Source\]](#). I built lightweight metadata csvs out of bulky HDF files using Spark and Python libraries. As a Software Associate at Goldman Sachs, I came across multiple dimensionality reduction techniques like Principal Component Analysis (PCA) and Recurring Feature Elimination (RFE), which we later used to compress data and create more accurate credit scoring models. As a Computer Science Graduate Student at UMass Amherst presently, I took Compression of Deep Neural Networks as my first-semester capstone project, where my team and I explored Quantization, Pruning, and Knowledge Distillation techniques to compress the ResNet-18 model.

Even though I have explored multiple facets of Compression, I have always been fascinated by Autoencoders and their flexibility to compress/decompress data in small latent spaces. This project is the perfect opportunity for me to get exposure on SOTA AutoEncoder frameworks like Baler and measure its performance while compressing ATLAS data. Additionally, Baler's robustness allows it to be tuned datasets outside of CERN unlocking oppurtunites for Baler to scale and become the go-to easy-to-use open-source lossy compression library. With this qualification task, I hope to prove my capabilities as a competent software engineer and my enthusiasm for Deep Neural Networks and Data Compression. To that end, I hope you would guide me into writing a thorough proposal that covers all task ideas and expected results with appropriate milestones and help me get selected and complete this program.