Virtual Conference on
Applications of Statistical Methods and Machine Learning
in the Space Sciences – 17-21 May 2021

Chris J. Maddison
University of Toronto, Canada
Lossy Compression for Lossless Prediction
Yann Dubois, Benjamin Bloem-Reddy, Karen Ullrich, Chris J. Maddison

Most data is automatically collected and only ever "seen" by algorithms. For example, thenumber of images in many sky surveys is so large that visual inspection is impractical. Despite this, classicallossy data compression methods are designed to store the information needed to guarantee perceptual fidelity rather than just the informationneeded byalgorithms performing downstream tasks. So, we are likely storing vast amounts of unneeded information. In this talk, we characterize theminimum bit-rates required to ensure high performance on all predictive tasks that are invariant under a set of transformations. Based on our theory,we designunsupervised objectives for designing deep-learning-based compression methods, and we show that it is possible to achieve dramatic rate savingson standard datasets, including Galaxy Zoo, without decreasing predictive performance.