ZooVera: Novel Application of Transfer Learning for Classifying Galaxy Morphologies in Simulated LSST Data

Vadivel, Cibi (School: Lake Zurich High School)

The Vera C. Rubin Observatory is currently under construction and is set to conduct the 10-year Legacy Survey of Space and Time (LSST) when complete. The LSST will catalog the entire night sky, capturing 20 billion galaxies and producing 60 petabytes of data. This will be thousands of times larger than any previous catalog, leaving astronomers the challenge of processing this data into an accurate representation of the universe. Machine Learning (ML) has been a useful tool in the past for analyzing sky survey data. A prime example is Zoobot, a Bayesian Convolutional Neural Network trained on millions of volunteer responses from Galaxy Zoo. This study retrained Zoobot via transfer learning to classify galaxy morphologies in simulated LSST data, producing a new model: ZooVera. Due to the nature of sky surveys, data captured in the first year (Y1) of the LSST will be much noisier than the final years. As a timely incentive, ZooVera was trained to distinguish between spiral, elliptical, and merging galaxies from mock Y1 data, which was extracted from the IllustrisTNG simulation with artificial Poisson noise. ZooVera produced performance metrics significantly higher than traditional ML models on the same class of data, with a much smaller training size. The success of ZooVera provides astronomers with a versatile tool that utilizes transfer learning to easily adapt to galaxy classification tasks on completely new data. When real LSST observations become available, crowdsourcing may allow large sets of galaxies to be labeled much faster than previously seen. ZooVera can also be trained entirely on Google Colab's free-to-use GPUs to enable novel scientific discoveries. This is a leap forward for transfer learning in the field of astronomy.