![PHYS.ORG]

Machine learning algorithms can find anomalous needles in cosmic haystacks

September 25 2023, by Scott Alan Johnston

Credit: ESA/Webb, NASA & CSA, J. Rigby

The face of astronomy is changing. Though narrow-field point-and-shoot astronomy still matters (JWST anyone?), large wide-field surveys promise to be the powerhouses of discovery in the coming decades, especially with the advent of machine learning.

A recently developed machine learning program, called ASTRONOMALY, scanned nearly four million galaxy images from the Dark Energy Camera Legacy Survey (DECaLS), discovering 1,635 anomalies including 18 previously unidentified sources with "highly unusual morphology." It is a sign of things to come: a partnership between humans and software that can do better observational science than either could do on their own.

Survey telescopes have long been part of the astronomers' toolkit. The difference in the 21st century is that they can now produce incredibly vast amounts of data, far more than a human could hope to dig through and examine on their own. The upcoming Vera Rubin Observatory, for example, is expected to create 20 terabytes of data every single night (60 petabytes over 10 years), and ultimately provide "32 trillion observations of 20 billion galaxies."

Poring through all that data would take humans decades. AI can do it much faster.

Most previous anomaly detection programs were trained on test datasets, teaching the algorithm to look for specific phenomena. The limitation of these programs is that they tend to find many anomalies of the same type, rather than entirely new anomalies.

ASTRONOMALY is instead run "unsupervised," allowing it to find new kinds of outliers—the kind of thing that gets astronomers excited, like gravitational lenses, galactic mergers, odd red-shift patterns, and anything else that is just weird. However, ASTRONOMALY performs best when it employs a form of active learning, with input from humans to correct its mistakes. Incorporating this feedback into its searches offers much better results.

The best part: it only takes the astronomer a few hours.

In a [recent preprint paper](link) posted to *arXiv*, astronomers tested ASTRONOMALY on a larger dataset than ever before, demonstrating that it can work at scale. After feeding the program a huge amount of DECaLS data, they tested several different algorithms. The results showed that the unsupervised method, enhanced by active learning input from humans, offered the highest output of unique anomalies.

The most interesting anomalies, according to the researchers, included "ring [galaxies](link) exhibiting strange colors and morphology, a source that is half red and half blue, a potential strongly lensed system with a pair of sources acting as the lens, several known interacting groups and some sources that are either interacting or coincidental alignments."

One puzzling object is giving off radio emissions that might be explained by the presence of a quasar, but the galaxy also has a ring feature that is either a rare red-ringed galaxy or a gravitational lens. Another anomaly looks to be a ring-shaped starburst galaxy with either a tidal tail or a colliding companion galaxy.

All of these rare objects would have been missed without the active learning algorithm. The results promise exciting new finds in the very near future.

But there is still one challenge to overcome in this new age of enormous datasets: data transfer.

"One of the main challenges that we experienced was the transfer of data from the host server to a local computer, which took several weeks," the researchers said. Their proposed solution? In the future, it makes more sense to bring the [computational power](link) to the host observatory, rather than try and bring the data offsite.

More information: Verlon Etsebeth et al, Astronomaly at Scale: