

April 23, 2018

Contact: Tim Stephens (831) 459-4352; stephens@ucsc.edu

Face recognition for galaxies: Artificial intelligence brings new tools to astronomy

A 'deep learning' algorithm trained on images from cosmological simulations has been surprisingly successful at classifying real galaxies in Hubble images

For Immediate Release

SANTA CRUZ, CA--A machine learning method called "deep learning," which has been widely used in face recognition and other image- and speech-recognition applications, has shown promise in helping astronomers analyze images of galaxies and understand how they form and evolve.

In a new study, accepted for publication in *Astrophysical Journal* and available online <https://arxiv.org/pdf/1804.07307.pdf>, researchers used computer simulations of galaxy formation to train a deep learning algorithm, which then proved surprisingly good at analyzing images of galaxies from the Hubble Space Telescope.

The researchers used output from the simulations to generate mock images of simulated galaxies as they would look in observations by the Hubble Space Telescope. The mock images were used to train the deep learning system to recognize three key phases of galaxy evolution previously identified in the simulations. The researchers then gave the system a large set of actual Hubble images to classify.

The results showed a remarkable level of consistency in the neural network's classifications of simulated and real galaxies.

"We were not expecting it to be all that successful. I'm amazed at how powerful this is," said coauthor Joel Primack, professor emeritus of physics and a member of the Santa Cruz Institute for Particle Physics (SCIPP) at UC Santa Cruz. "We know the simulations have limitations, so we don't want to make too strong a claim. But we don't think this is just a lucky fluke."

Galaxies are complex phenomena, changing their appearance as they evolve over billions of years, and images of galaxies can provide only snapshots in time. Astronomers can look deeper into the universe and thereby "back in time" to see earlier galaxies (because of the time it takes light to travel cosmic distances), but following the evolution of an individual galaxy over time is only possible in simulations. Comparing simulated galaxies to observed galaxies can reveal important details of the actual galaxies and their likely histories.

In the new study, the researchers were particularly interested in a phenomenon seen in the simulations early in the evolution of gas-rich galaxies, when big flows of gas into the center of a galaxy fuel formation of a small, dense, star-forming region called a "blue nugget." (Young, hot stars emit short "blue" wavelengths of light, so blue indicates a galaxy with active star formation, whereas older, cooler stars emit more "red" light.)

In both simulated and observational data, the computer program found that the "blue nugget" phase only occurs in galaxies with masses within a certain range. This is followed by quenching of star formation in the central region, leading to a compact "red nugget" phase. The consistency of the mass range was an exciting finding, because it suggests the deep learning algorithm is identifying on its own a pattern that results from a key physical process happening in real galaxies.

"It may be that in a certain size range, galaxies have just the right mass for this physical process to occur," said coauthor David Koo, professor emeritus of astronomy and astrophysics at UC Santa Cruz.

The researchers used state-of-the-art galaxy simulations (the VELA simulations) developed by Primack and an international team of collaborators, including Daniel Ceverino (University of Heidelberg), who ran the simulations, and Avishai Dekel (Hebrew University), who led analysis and interpretation of them and developed new physical concepts based on them. All such simulations are limited, however, in their ability to capture the complex physics of galaxy formation.

In particular, the simulations used in this study did not include feedback from active galactic nuclei (injection of energy from radiation as gas is accreted by a central supermassive black hole). Many astronomers consider this process to be an important factor regulating star formation in galaxies. Nevertheless, observations of distant, young galaxies appear to show evidence of the phenomena leading to the blue nugget phase seen in the simulations.

For the observational data, the team used images of galaxies obtained through the CANDELS project (Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey), the largest project in the history of the Hubble Space Telescope. First author Marc Huertas-Company, an astronomer at the Paris Observatory and Paris Diderot University, had already done pioneering work applying deep learning methods to galaxy classifications using publicly available CANDELS data.

Koo, a CANDELS co-investigator, invited Huertas-Company to visit UC Santa Cruz to continue this work. Google has provided support for their work on deep learning in astronomy through gifts of research funds to Koo and Primack, allowing Huertas-Company to spend the past two summers in Santa Cruz, with plans for another visit in the summer of 2018.

"This project was just one of several ideas we had," Koo said. "We wanted to pick a process that theorists can define clearly based on the simulations, and that has something to do with how a galaxy looks, then have the deep learning algorithm look for it in the observations. We're just beginning to explore this new way of doing research. It's a new way of melding theory and observations."

For years, Primack has been working closely with Koo and other astronomers at UC Santa Cruz to compare his team's simulations of galaxy formation and evolution with the CANDELS observations. "The VELA simulations have had a lot of success in terms of helping us understand the CANDELS observations," Primack said. "Nobody has perfect simulations, though. As we continue this work, we will keep developing better simulations."

According to Koo, deep learning has the potential to reveal aspects of the observational data that humans can't see. The downside is that the algorithm is like a "black box," so it is hard to know what features in the data the machine is using to make its classifications. Network interrogation techniques can identify which pixels in an image contributed most to the classification, however, and the researchers tested one such method on their network.

"Deep learning looks for patterns, and the machine can see patterns that are so complex that we humans don't see them," Koo said. "We want to do a lot more testing of this approach, but in this proof-of-concept study, the machine seemed to successfully find in the data the different stages of galaxy evolution identified in the simulations."

In the future, he said, astronomers will have much more observational data to analyze as a result of large survey projects and new telescopes such as the Large Synoptic Survey Telescope, the James Webb Space Telescope, and the Wide-Field Infrared Survey Telescope. Deep learning and other machine learning methods could be powerful tools for making sense of these massive datasets.

"This is the beginning of a very exciting time for using advanced artificial intelligence in astronomy," Koo said.

In addition to Primack, Koo, and Huertas-Company, the coauthors of the paper include Avishai Dekel at Hebrew University in Jerusalem (and a visiting researcher at UC Santa Cruz); Sharon Lapiner at Hebrew University; Daniel Ceverino at University of Heidelberg; Raymond Simons at Johns Hopkins University; Gregory Snyder at Space Telescope Science Institute; Mariangela Bernardi and H. Dominquez Sanchez at University of Pennsylvania; Zhu Chen at Shanghai Normal University; Christoph Lee at UC Santa Cruz; and Berta Margalef-Bentabol and Diego Tuccillo at the Paris Observatory.

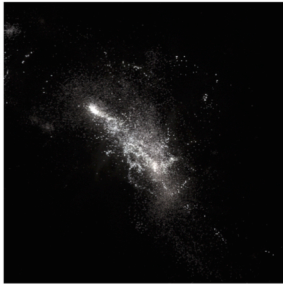
This work was partly supported by grants from France-Israel PICS, US-Israel Binational Science Foundation, U.S. National Science Foundation, and Hubble Space Telescope. The VELA computer simulations were run on NASA's Pleiades supercomputer and at DOE's National Energy Research Scientific Computer Center (NERSC).

#####

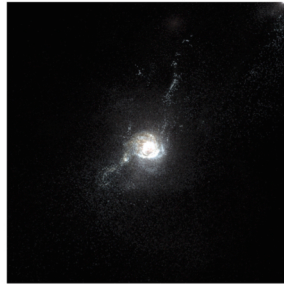
Note to reporters: You may contact Joel Primack at joel@ucsc.edu (cell: 831-345-8960) and David Koo at koo@ucolick.org.

For more UCSC news, visit news.ucsc.edu.

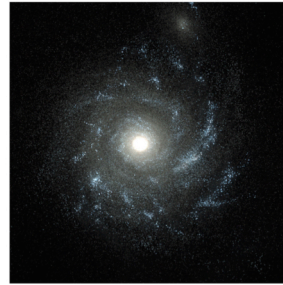
Pre-Blue-Nugget-Stage



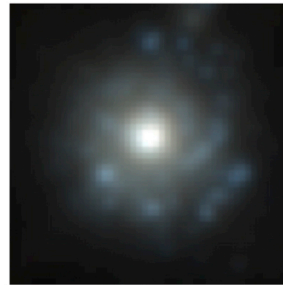
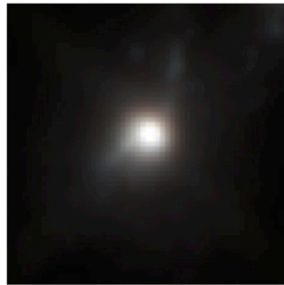
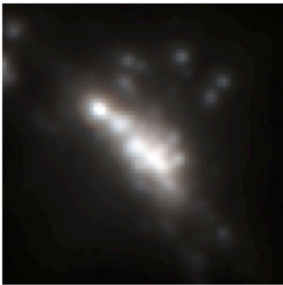
Blue-Nugget-Stage



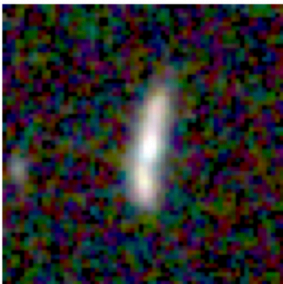
Post-Blue-Nugget-Stage



High-resolution images from a computer simulation of a young galaxy going through the 3 phases of evolution



Same images from the computer simulation of a young galaxy going through the 3 phases of evolution, as it would be observed by Hubble Space Telescope



Hubble Space Telescope images of distant young galaxies classified into the 3 phases with a deep learning algorithm

Examples of simulated galaxy evolution through the 3 stages: Pre-Blue Nugget (often elongated, i.e. pickle shaped), Blue Nugget (compaction phenomenon: gas infall leads to central starburst), and Post-Blue Nugget (often with star-forming disk), with similar galaxies observed by Hubble Space Telescope. The width of each image is approximately 100,000 light years. Credits: simulations Daniel Ceverino and Joel Primack, simulated images Greg Snyder and Marc Huertas-Company, HST observations CANDELS.