Finding similar galaxies using deep learning methods

Edukov Mikhail
mike.edukov@gmail.com

Scientific supervisor: Dr. Savchenko S.S., Department of Astrophysics, Saint-Petersburg State University

We present the results of the application of automated clustering methods to a sample of galaxies images. The main goal is to develop a tool for finding galaxies, that are similar to a given one. Our purpose is to investigate methods of selecting and extracting comparable features of galaxy images. To achieve this goal we applied deep learning, in particular, autoencoder techniques.

Autoencoder is an unsupervised learning algorithm where a neural network is trained to copy its input to its output. In other words, it is trying to learn to approximate identity function, but with some constraints on the network, such as decreasing of the size of hidden layers, making it significantly less than input size, so we can reveal common patterns in data. Learning to reconstruct the input image from incomplete representation forces the autoencoder to detect the most important features of the data. For our problem, we are only interested in the hidden layer, which is represented by a vector. Using it as a marker of galaxy structure, shape, and other visual properties and comparing the measure of this vector we can reason about the similarity with other images.

We have applied this method to a sample of galaxies from Sloan Digital Sky Survey Data Release 14. On this stage of research, we selected galaxies with perceptible structure and large enough visible radius (around 7.5k images in total). All these images were used in training with augmentation (synthetically modified data) for dataset expansion. The current version of our software computes vector representation of galaxy and compares it with all other in demo database. It is worth noting that there are no robust methods that calculate likeness of content on two images except comparing by eyes. But our approach allows one to get an approximately similar galaxy, given properties such as class, shape, color, and other distinctive features. In the future, we plan to use another deep learning techniques and architecture to improve extraction and getting more explainable features. Also, this network can be applied and trained on a larger set of images.