The ability to get a quick match of a galaxy image via a simple web application could help streamline the data collection process for astronomers.

Using Deep Learning with a Web API to Match Galaxy Images

Lauren Gregory and Matias Carrasco Kind

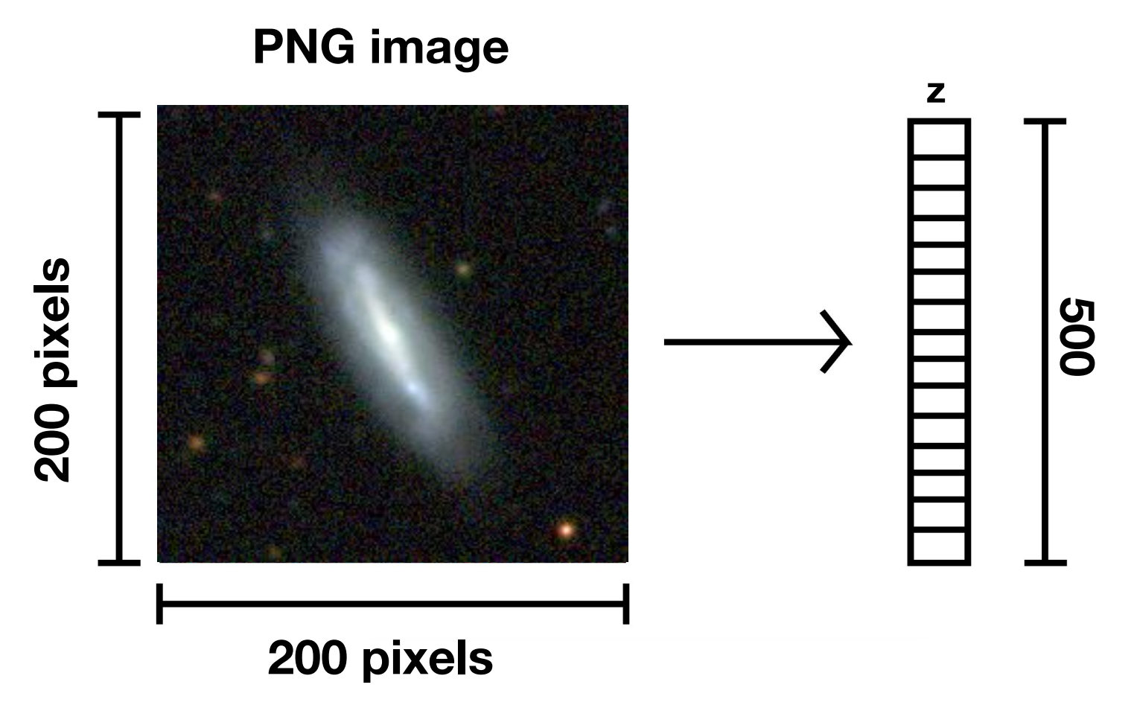
Department of Astronomy and the National Center for Supercomputing Applications

Motivation

The Dark Energy Survey (DES) is an international project that aims to map hundreds of millions of galaxies and find patterns in cosmic structure to learn more about dark energy. Currently, the DES catalog hosts more than 700 million objects, with around 500 million of them being galaxies and most of those being faint galaxies. Finding similar looking galaxies can be hard to do by eye, and even harder when you have such a high volume of images to search through and compare. The DES Interface allows users to find observational data on thousands of objects, and that framework could be helped by a service that finds visual matches. Astronomers studying galaxy formation and evolution can find data related to similarly shaped galaxies to one they are studying to explore further relationships more efficiently than before, and inputting anomalous images could assist in finding objects where data is missing.

The Model

The model vectorizes images using an autoencoder from Keras, a deep learning application that uses TensorFlow to construct neural networks. This compresses the image into a vector which can then be compared to a database of other image vectors that have been previously compressed by Keras. This database is in place so that the model doesn't have to run on every galaxy each time a request is run. The euclidean distance between the main vector and all of the vectors in the database is used to determine which galaxies are the closest in appearance to the original.



The galaxy image and it's vectorized counterpart

$$d = \sqrt{\sum_{i=1}^{N} (Xi - Yi)^2}$$

The Euclidean distance formula between image vectors X and Y

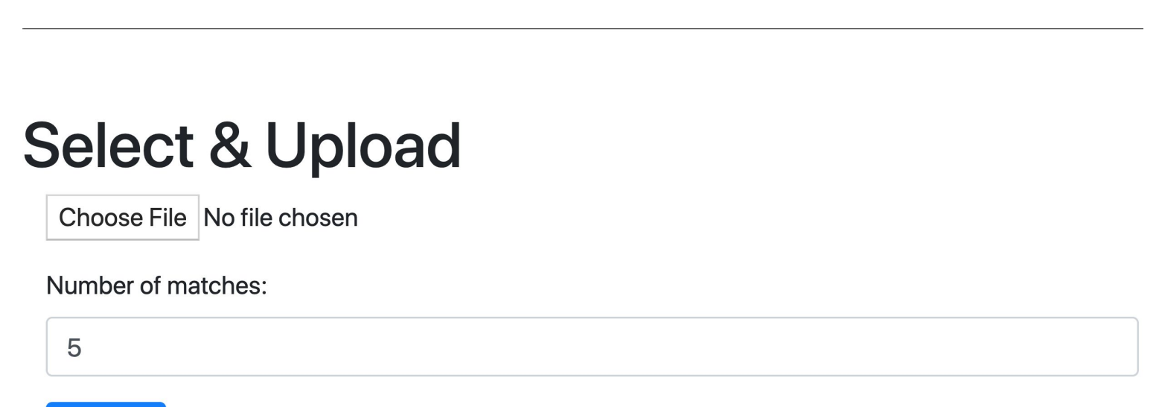
The API

The API is written in Python and uses the Python web framework Tornado. It starts out with two user inputs in a simple html form. These inputs are the image (.jpg or .png) and *n*, the number of matching galaxy images desired.

When the form is submitted, the post request of the API runs the model to compress the input image. It then compares it to the database of compressed images and computes the euclidean distance between the vector from the input image and the thousands of vectors in the loaded database.

When the distances are found, the lowest *n* are chosen and their filenames are fetched from a separate file. Then, using these filenames and Tornado's templating feature, a .html file is generated off a template that shows the matching images. It appears under the form, so the page doesn't need to be refreshed when another image match is needed.

This app has been containerized through Docker so that it and all of its dependencies can be seamlessly added to part of the DES interface or used locally with minimal installations needed.



Select & Upload

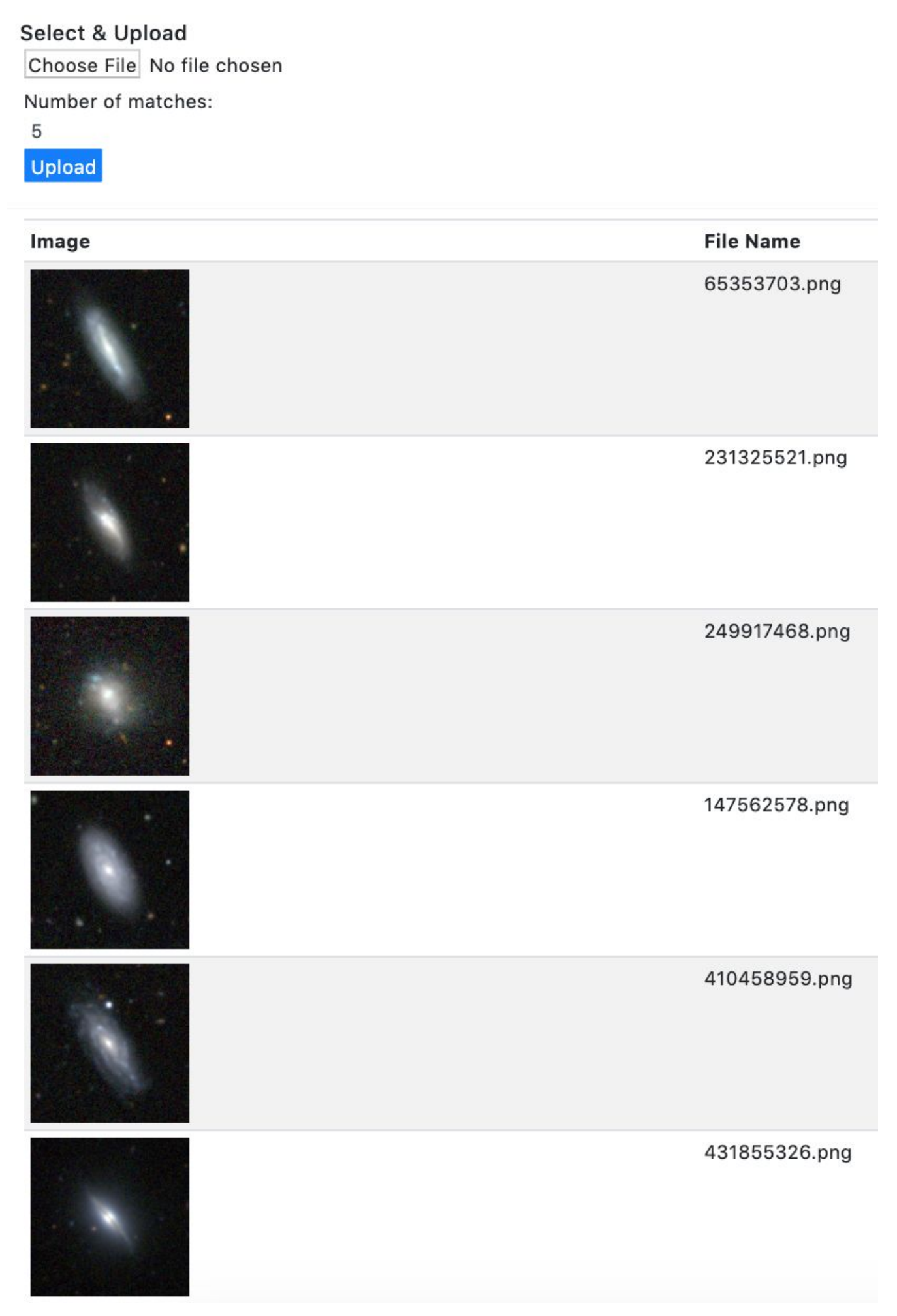
Choose File No file chosen

Number of matches:

5

Upload

The upload form, with 5 as the default n value



Select & Upload

Choose File No file chosen

Number of matches:

5

Upload

Image	File Name
	65353703.png
	231325521.png
	249917468.png
	147562578.png
	410458959.png
	431855326.png

The result of the API, n=5 matching galaxy images

Future Directions

This project is a great start to a larger and more comprehensive service for DES researchers. Steps that can be made to increase the utility of this app include supplemental data output for each galaxy image from the database and expanding to a larger dataset.

Returning accompanying data with the image would streamline the process of data collection for researchers, especially if it was made available in a downloadable .csv file format. They could input an interesting galaxy they are studying and instantly have access to important data from galaxies that are possibly in a similar stage of their evolution, or anomalous galaxies that may have characteristics in common.

Additionally, expanding the dataset will allow for more accurate matches due to a larger pool to draw from. When the dataset is expanded, the model currently in use will give way to multiple smaller models that will each operate on sets of objects binned by magnitude. This will ensure that the most accurate matches are being served back to the user.

NSF DARK ENERGY SURVEY NCSA

The DES Data Management System is supported by the National Science Foundation under Grant NSF AST 07-15036 and NSF AST 08-13543