

Windows On Earth

Aditi Dass, Greg Frasco, Frederick Joossens
aditid@bu.edu, frascog@bu.edu, fjansen@bu.edu

Task

In this project we will be classifying images that are taken from the International Space Station to be categorized into galleries on <http://www.windowsonearth.org/>. The set of tags will be coming from the Windows of Earth organization. One of the difficulties of this project is that only a small subset of the images is labeled, and the total size of the dataset is over 1.3 million images. See Fig 1 for example image and tags.

Data Collection and Processing

Aside from the sample data set provided, we plan to write a script to download the set of all tagged images for supervised learning. This will be split into subsets for training, verification and testing. Another randomly chosen set of downloaded images will be collected for unsupervised learning data set.

Preprocessing will involve cleaning the data set, downsizing the image resolution and experimenting with filters such as grayscale, sobel, average colour values and regions of interest.

Further research will be done on possible feature extraction beyond the techniques taught in class.

Approach

Our approach is to create several different machine learning models, supervised and unsupervised, for each image classification tag. While we are going to use clustering to identify key features in different photo environments, we also intend to start with a basic neural network that would lead to the

binary classification (logistic regression) of whether an image fits a tag or not. Further experimentation with other models such as decision tree and random forests may be experimented with to further refine our results.

Milestones

Oct 25th: Infrastructure set up on SCC, example PyTorch scripts and data processing pipeline

Oct 31st: Research Unsupervised models for image classification

Nov 7th: Build Neural Network infrastructure and supervised models

Nov 30th: Train / validate final models

Dec 7th: Final Report Completion

Dec 15th: Project Poster Completion

Member Roles

Aditi Dass: Unsupervised model experimentation + supervised models

Greg Frasco: Supervised models

Frederick Joossens: Data collection + Infrastructure pipeline + Preprocessing



Figure 1: Example Instance
Tags: Dragon, Dock Undock, Structure