Windows On Earth

Aditi Dass, Greg Frasco, Frederick Joossens

CS 542: Machine Learning

BOSTON UNIVERSITY

Introduction

In this project we will be classifying images that are taken from the International Space Station to be categorized into galleries on www.windowsonearth.org. The set of tags will be coming from the Windows on 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.

Approach

Our approach is to create several different machine learning models, supervised and unsupervised, for each image classification tag. We tested various clustering methods to identify key features in different photo environments, as well as a basic neural network that would lead to the binary classification (logistic regression) of whether an image fits a tag or not, and finally transfer learning to rely on a pre-trained model.

Figure 1:
Tags:
Dock Undock
Spacecraft
Dragon
ISS Structure
Day
Clouds

Classification Methods

Supervised Learning

Each supervised learning classifier was trained on the initial set of 10000 photos that was provided, and completed a five cross-fold validation on a set of classifiers for each tag to get a mean accuracy and a standard deviation. We choose

- Many standard classifiers for image classification such as kNN, random forests, and SVMs
- Testing on networks that are not usually used for image classification such as Naive Bayes

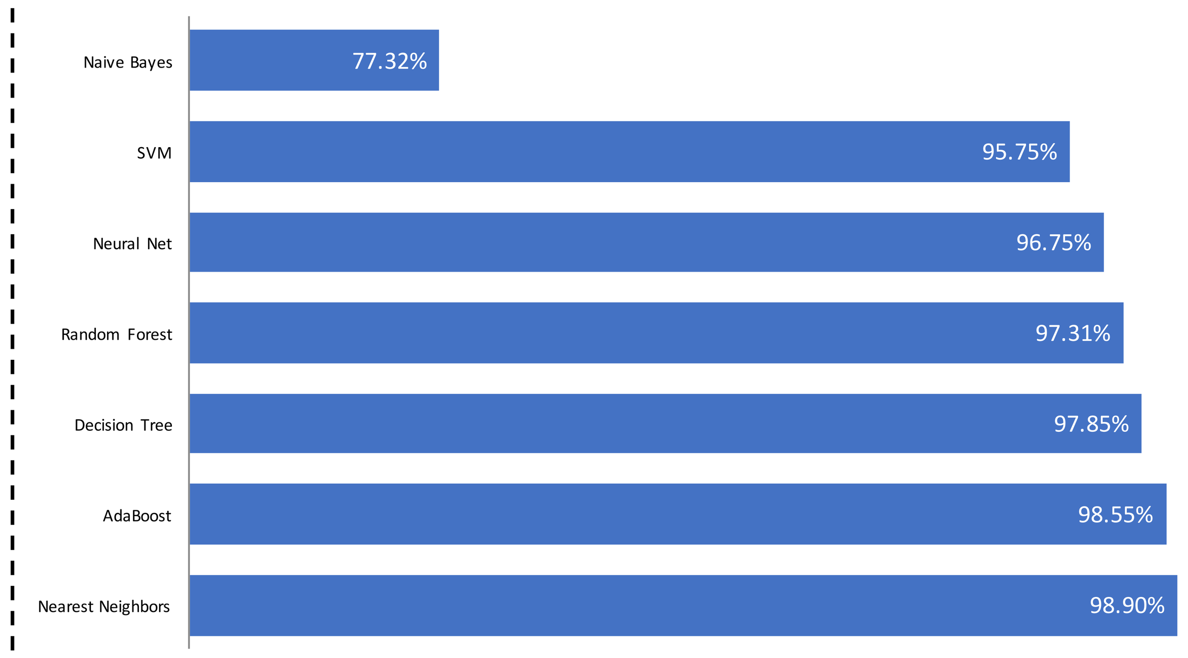

Figure 2: Accuracy vs Classifier

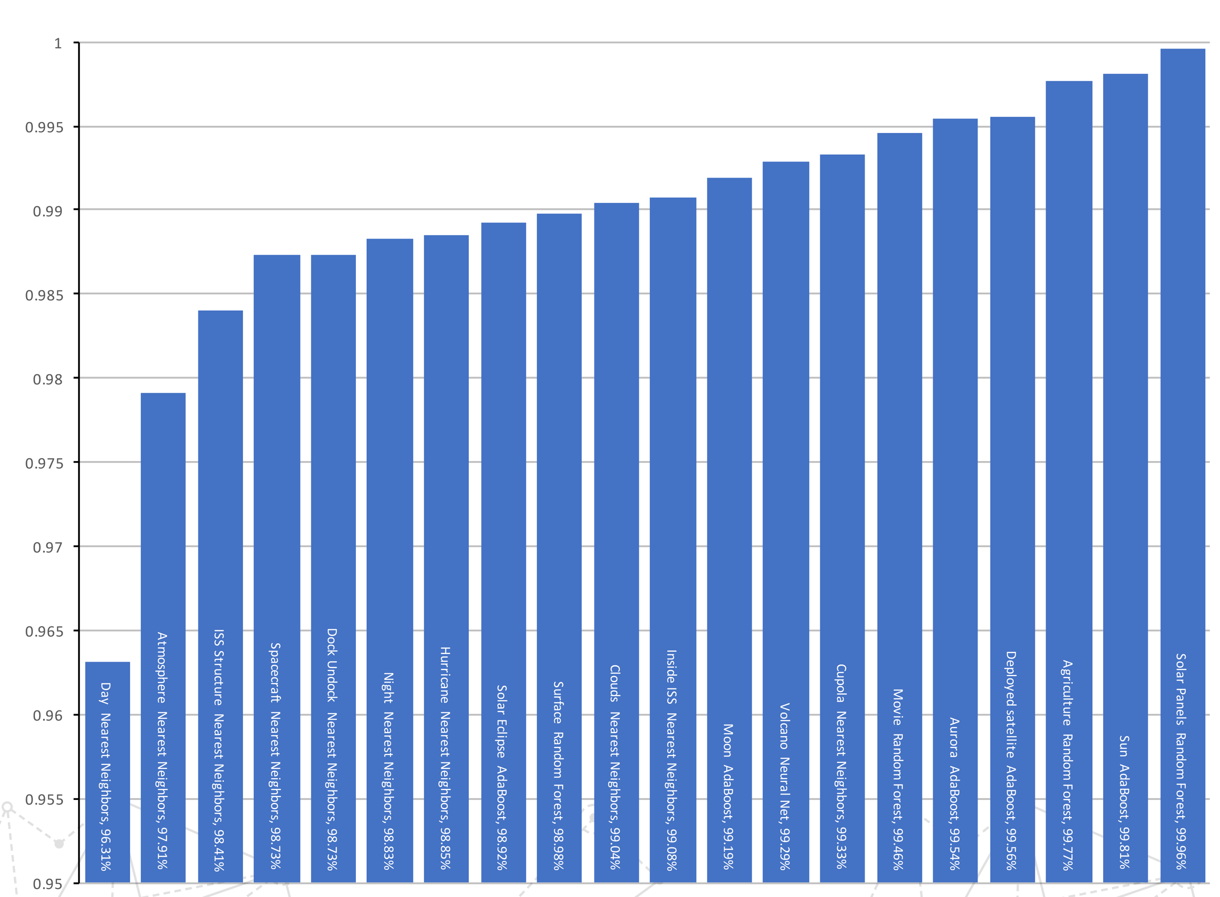

Figure 3: Accuracy vs Tag

Movie Series

To identify a movie-series we utilized a number of comparison methods. However, the data set provided did not have a consistent provision for the tag "movie" that fit the definition provided. Therefore, we used a variety similarity metrics on both the tagged images and a sample of random images. By comparing these sets with a two sample T test, we were able to determine if the similarity metric was statistically significant enough to act as an identifier of a movie series.

Method	P-values
Mean Square Error	1.2349×10^{-29}
Naive Entropy	5.9614×10^{-57}
PT Entropy	5.0377×10^{-68}
QE Entropy	2.0626×10^{-66}
Image-Histogram Entropy	**6.425×10^{-76}**

Figure 4: Similarity Metric Results

Values for successive images provided by the above had extremely small p-values which means we were able to reject the premise that the difference between the similarity in random images and similarity in time-lapsed images was random with a very high confidence interval.

Transfer Learning

For transfer learning we use the pre-trained Inception-v3 model and retrain the final fully connected layer. In order to support single label classification as well as multi-label classification, we attempted three different approaches:

- Utilizing softmax for the activation function and a softmax cross entropy loss. This matches the approach that Inception-v3 takes to train the model.
- Changing the last layer to a sigmoid nonlinearity with a sigmoid cross entropy loss in order to support multiple labels. This matches a technique employed by Google for the Open Images dataset.
- Using the single-label classification technique to obtain a multi-label classification result by training a separate, binary model for each individual label. This allows us to use a softmax activation function instead of sigmoid, which makes fine-tuning the data set per label easier.

Figure 5 Multi-Label Validation Accuracy for Transfer Learning