# Classification of TERC Photos: A Space Odyssey

Dimitri Makrigiorgos, Neela Kaushik, Sarah Ferry

*Data Collection/Processing*

Data collection will involve downloading the training dataset and retrieving the classification labels from the image metadata. Preprocessing on the images will involve vectorizing and reducing the size of each image in order to allow for them to be processed by the neural network.

*Plan for Approach*

We will use a convolutional neural network to analyze the images and output classifications based on pixels. The model will be trained on images for which TERC has already provided labels. The input to the model is an array of pixel values corresponding to an image, on which we will use a Cross-entropy binary classifier for each potential tag. The output of the model is the a binary classification of 1 or 0 for each of the tags. We will use Python along with the TensorFlow and/or scikit-learn libraries to implement the CNN and classification algorithms.

To detect the movies, we will use OpenCV libraries to detect image similarity. If images are similar to within some certainty (e.g. 95%), their image IDs are consecutive and/or their timestamps are within one second of each other indicating that they were taken consecutively, then the set of images will be classified as a movie. The images will still be tagged with the output of the CNN classifier.

The final product will be a program which takes as input a list of image file names and returns a CSV which contains the image name associated with the tags defined by the convolutional neural network.

*Milestones*

1. **Tuesday, October 24th:** Project proposal due.
2. **Monday, October 30th:** Gather requirements, research CNNs and develop a more comprehensive plan.
3. **Monday November 21st:** Initial coding complete, submit initial results.
4. **Monday December 7th:** Project complete, submit write-up final results.
5. **Tuesday December 12th:** Project and poster complete.

*Team Member Roles*

1. Project Point Person: Dimitri
2. Programming: Neela, Sarah, Dimitri

*Approval by Project Partner David Libby received 10/23/2017*