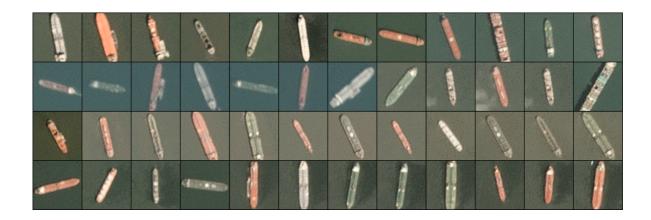# Ships in Satellite Imagery Dataset

Context

- Satellite imagery provides unique insights into various markets, including agriculture, defense and intelligence, energy, and finance.
- New commercial imagery providers, such as Planet, are using constellations of small satellites to capture images of the entire Earth every day.
- This flood of new imagery is outgrowing the ability for organizations to manually look at each image that gets captured, and there is a need for machine learning and computer vision algorithms to help automate the analysis process.
- The aim of this dataset is to help address the difficult task of detecting the location of large ships in satellite images.
- Automating this process can be applied to many issues including monitoring port activity levels and supply chain analysis.

Content

- The dataset consists of image chips extracted from Planet satellite imagery collected over the San Francisco Bay and San Pedro Bay areas of California.

- It includes 4000 80x80 RGB images labeled with either a "ship" or "no-ship" classification.

- Image chips were derived from PlanetScope full-frame visual scene products, which are orthorectified to a 3 meter pixel size.

- Provided is a zipped directory `ship_images.zip` that contains the entire dataset as .png image chips.

- The images are divided into two sub-folders: **ship** and **no-ship**.

- Each individual image filename follows a specific format: {label} __ {scene id} __ {longitude} _ {latitude}.png

  - **label:** Valued 1 or 0, representing the "ship" class and "no-ship" class, respectively.

  - **scene id:** The unique identifier of the PlanetScope visual scene the image chip was extracted from.

    - The scene id can be used with the Planet API to discover and download the entire scene.
  - **longitude_latitude:** The longitude and latitude coordinates of the image center point, with values separated by a single underscore.

Class Labels

- The "ship" class includes 1000 images.
- Images in this class are near-centered on the body of a single ship.
- Ships of different sizes, orientations, and atmospheric collection conditions are included.
- Example images from this class are shown below.

- The "no-ship" class includes 3000 images.
- A third of these are a random sampling of different landcover features - water, vegetation, bare earth, buildings, etc. - that do not include any portion of an ship.
- The next third are "partial ships" that contain only a portion of an ship, but not enough to meet the full definition of the "ship" class.
- The last third are images that have previously been mislabeled by machine learning models, typically caused by bright pixels or strong linear features.
- Example images from this class are shown below.

Objective:

- This is an Binary Image Classification dataset.
- It can be used to train a model capable of differentiating images containing ships and those not containing ships.
- Transfer Learning can be applied while training the model.