# SimCLR Final Report

**By Aidan Lee-Wen and Jeremy Mumford**

**Abstract**

The goal is to apply Google's SimCLR model to satellite imagery. Full-sized images were used to train but failed to produce accurate results. RinCon's custom chipped images were also used, but contained multiple labels per image and noise from terrain. It is recommended that training is performed on images of each predetermined object of interest and then identified as being present in full-sized satellite images. The team believes that this recommended process will help remove the noise from terrain and align with realistic satellite imagery.

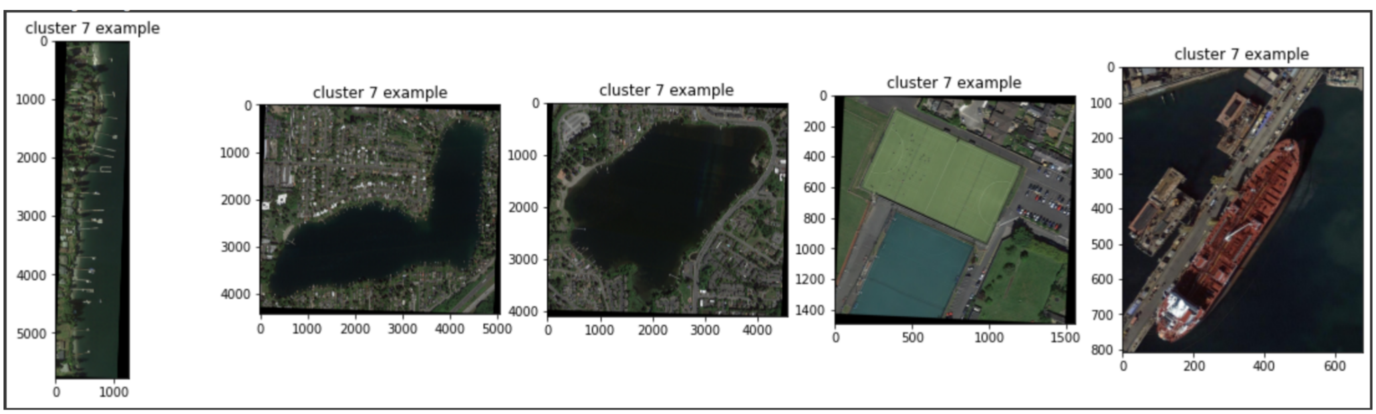
**Methods**

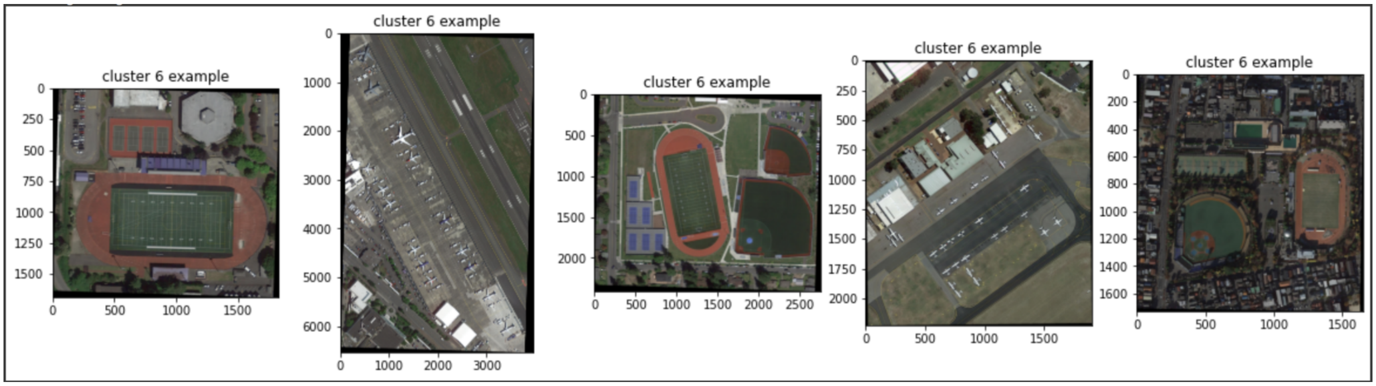
In 2020, Google had released SimCLR: a new, groundbreaking unsupervised contrastive learning model. The identified task was to learn about SimCLR and apply it to the DOTA dataset, which consists of satellite images.

While Google released their code to the public, it is organized in a way that works seamlessly with training on Google Cloud with predetermined datasets. Since the desired dataset was not included, packages from the company 'Lightly' were discovered and used to curate training on the DOTA dataset. The sponsor company, Rincon Research Corporation, provided a custom dataset that consisted of chipped and labeled images of the original DOTA dataset of full-sized images. Training was initially run on the full-sized images, but failed to produce accurate results. This resulted in a shift in attention to the chipped images.

**Results**

Training run on the full-sized images failed to produce accurate results. Examples can be found below:






Because the inaccurate results caused difficulties in discretionarily organizing the clusters with labels, there are no accuracy numbers. The inaccuracies are hypothesized to derive from two factors:

- Zoomed out, low resolution images makes it difficult for an algorithm to determine the focal object of the image
- One image can include multiple objects

While a human may look at these images and identify the focal object(s), such as planes or football fields, computer vision does not inherently distinguish between a zoomed-out airplane from the large runway it stands on (such as in the second cluster 6 example above).

**Recommendations**

The very poor results of the full-sized images provided insight on the struggles of satellite imagery. It is recommended that each object of importance is pre-specified, separated, and labeled. Then on the full satellite image, identify the focal objects of importance. For example, smaller, complete pictures of airplanes, cars, etc should be labeled and trained. Then on a full satellite image, identify that multiple planes and cars are present. The team's intuition believes that this should help remove noise from terrain and allow the model to separate terrain from important objects. This is also more aligned with reality, as satellite images are not going to be perfectly divided. For example, satellite images may not be separated into an airport, then another image is a football field. Rather, one image may be an airport with some airplanes but also include a football field beside it. The focal images of importance can then be identified as being present in the image: airplanes and a football field.

While training was performed on the chipped images, they do not separate objects of importance. Rather, they are cropped in a seemingly random manner. As a result, multiple objects are often included in one image or the image includes a half an object, half terrain. Some images in the chipped dataset are of nothing. This resulted in training being combined with noisy terrain and multiple labels being correct for one chipped image. Instead, it is recommended that the method suggested in the previous paragraph is performed. The team believes that this recommendation will help solve the difficulties of satellite imagery.

**Conclusion**

Applying a contrastive model on satellite imagery is difficult, especially when trying to identify the important characteristics of an image from noise from terrain and other objects. The team believes that the recommendation provided will help solve the issues associated with satellite imagery.