Reducing Effects of Swath Gaps in Unsupervised Machine Learning

Sarah Chen* **Esther Cao***
sarahc2@andrew.cmu.edu, esthercao888@gmail.com

Anirudh Koul Satyarth Praveen Meher Anand Kasam Siddha Ganju

Abstract

NASA Terra and NASA Aqua satellites capture Worldview data containing "swath gaps" which are "no data areas". Swath gaps can overlap the region of interest (ROI) completely, often rendering the entire imagery unusable by ML models. This problem is further exacerbated when the object rarely occurs (e.g. a hurricane) and, on occurrence, is partially overlapped with a swath gap. With annotated data as supervision, a model can learn to differentiate between the area of focus vs the swath gap. However, annotation is expensive and currently 35 PB of unannotated satellite data exists.

Our work builds on an observation by Seeley et al [1] - post unsupervised training of CNNs, outputs of the convolutional layers (activation maps) focus more on the swath gaps rather than the ROI for images where the swath occupies less than 25% area per image.

Our hypothesis is that reducing the detectable pattern or removing the existence of the gaps would allow the CNN to focus on the ROI. To prove this, we fill the swath gap with new pixel values, train a convolutional autoencoder, and then verify the classifier accuracy of nearest neighbors. We test our hypothesis on the UC Merced Land Use Dataset. The dataset originally lacked swath gaps, so we added gaps through empty polygons (up to 25% areas) and filled the swaths with three strategies: (1) random RGB pixel values sampled from a uniform distribution (2) randomly chosen non-null pixels from other parts of the same image (3) non-null neighboring pixels with probability inversely proportional to distance, i.e. closer points get chosen more often.

After training with a convolutional autoencoder, we retrieved the nearest neighbor images for images with the swath gaps filled and compared whether the retrieved category was the same as the query image. Images with swath gaps retrieved other images with swath gaps, giving us a useful baseline to improve on. Strategies 1 and 2 provided minor improvements to the nearest neighbors retrieval accuracy. On the other hand strategy 3 gave the exact same accuracy as the original swathless image (best case performance).

Additionally, the activation maps showed that the network does not pay attention to swath gaps. In certain cases the filled swath gaps looked so realistic that even a human evaluator could not distinguish between original satellite images and the swath gap filled images with strategy 3. Since our method is aimed at unlabeled data, it is widely generalizable and can potentially impact large scale unannotated datasets from space data domains.

*Equal Contribution; Work done as researchers at Space ML

Submitted to COSPAR 2021 Cross-Disciplinary Workshop on Machine Learning for Space Sciences, Sydney, Australia.

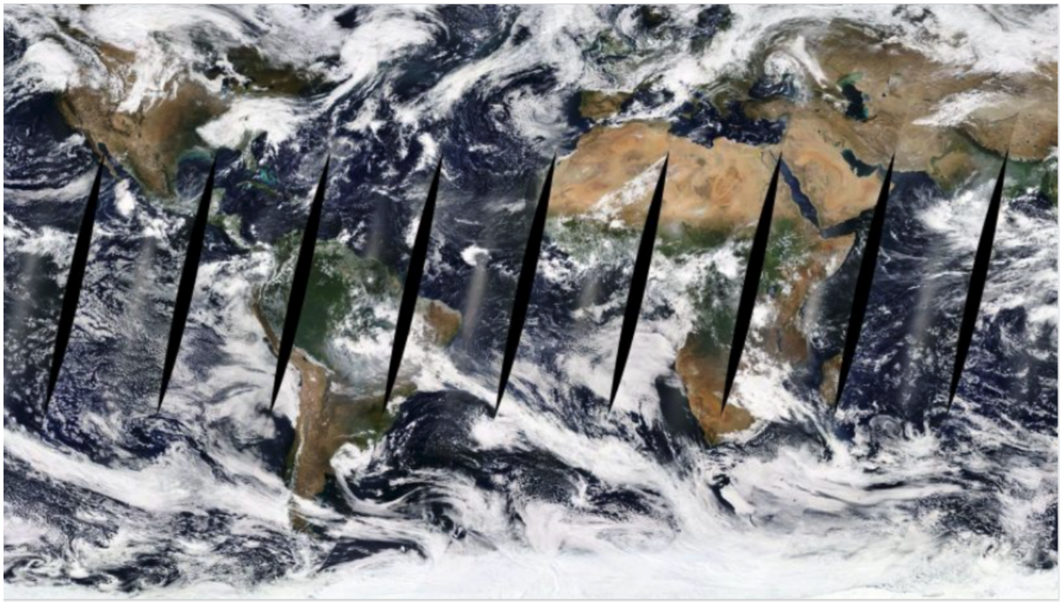
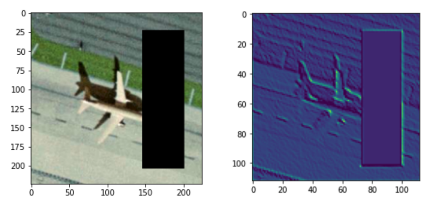
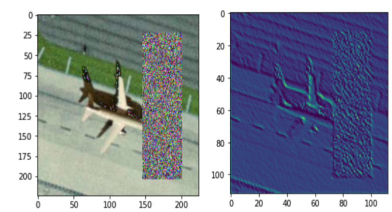


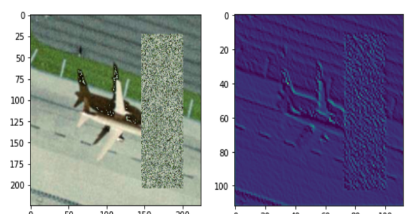
Figure 1: Swaths gaps on NASA Worldview which visualizes data from NASA Terra and NASA Aqua satellites.



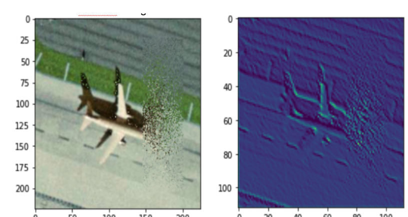
(a) Swath image and corresponding activation map



(b) **Strategy 1:** Swaths filled with pixels from random distribution and corresponding activation map



(a) **Strategy 2:** Swaths with pixels from distribution of nonempty pixels and corresponding activation map



(b) **Strategy 3:** Swaths filled with pixels from weighted distribution of neighboring pixels and corresponding activation map

Figure 2: Non-filled swath and three filling methods with activation maps

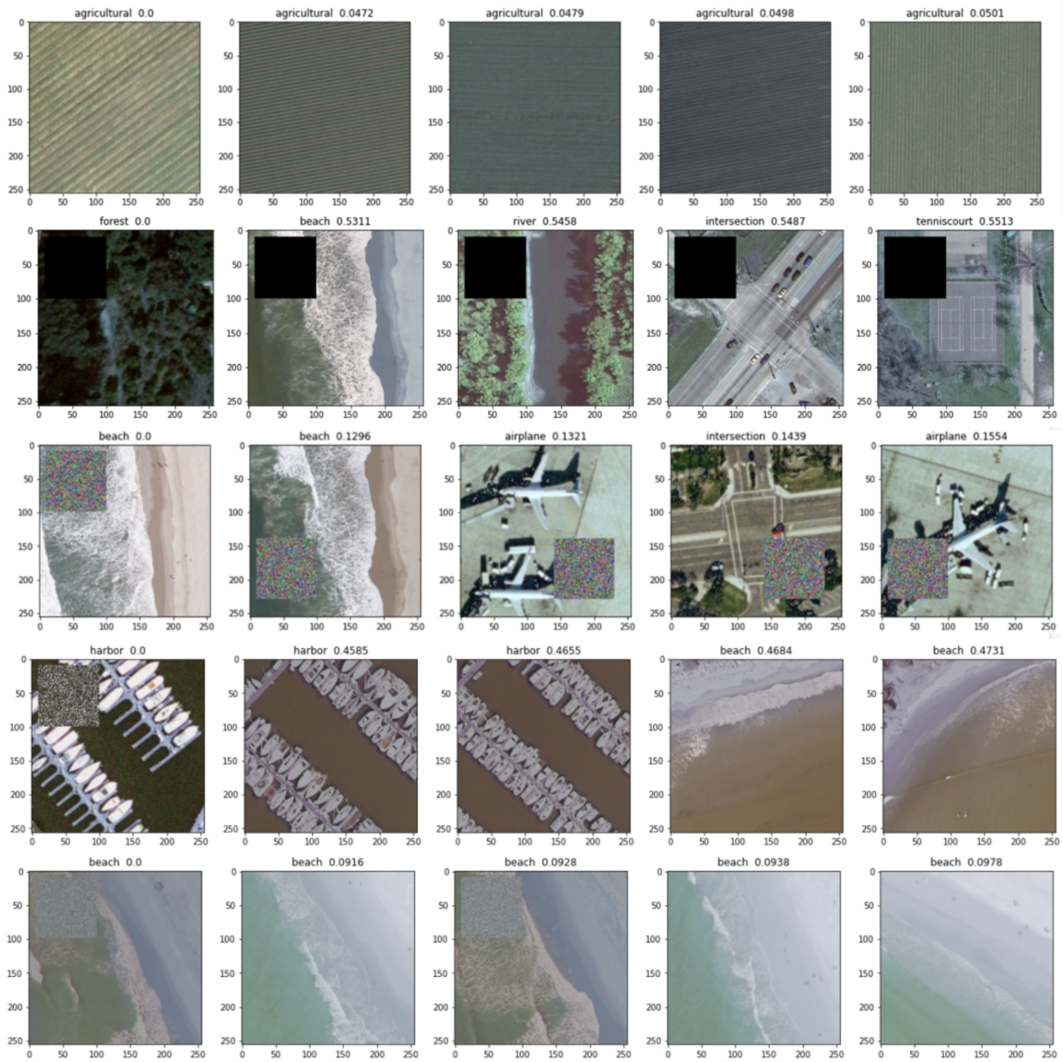


Figure 3: Trained convolutional autoencoder output. Query image (leftmost column) and the nearest four retrieved images through Nearest Neighbors. Swath filling strategy changes row wise (rows 1 - 5): no fill, Strategy 1, Strategy 2, Strategy 3. Strategy 1 filled swaths shows that the autoencoder learns the swath position. Strategy 3 shows that the trained autoencoder ignores the swath, concentrates on the region of interest thereby learning the features of interest.

References

[1] Seeley M., Civilini F., Srishankar N., Praveen S., Koul A., Berea A., El-Askary H. (2020) Knowledge Discovery Framework, NASA Frontier Development Lab, Available Online.