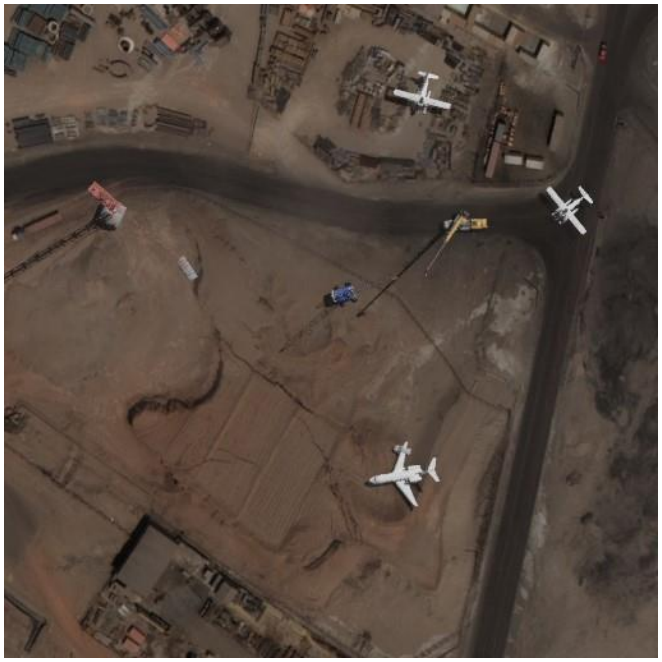
Orbital Insight

Orbital Insight Technical Report

Using Synthetic Imagery to Train Detection Models of Rare Objects in Satellite Imagery

July 2021



Point of Contact: Joe Weber
jweber@orbitalinsight.com

Overview/Executive Summary

The use of supervised deep learning has revolutionized the field of computer vision (CV). Deep learning models have achieved near-human-level performance across a broad range of CV tasks, including object detection, classification, semantic segmentation, and image generation. These models use supervised learning to train large networks of parameters, such as those used in convolutional neural networks (CNNs), on datasets that demonstrate how a task should be performed. But supervised learning comes with a significant cost — a critical need for large amounts of labeled data to train these models. At present, state-of-the-art algorithms are not able to learn from unlabeled data; they require large amounts of painstakingly labeled data in order to perform well, creating a labeled data-dependency problem. In the best case, this problem can be solved by manually labeling data, although that is time-consuming, expensive, and error prone. In the worst case, which is the situation with uncommon objects, dataset curation is even more difficult, as there isn't enough available data to label.

On the other hand, advancements in computer graphics have made it possible to generate a virtually unlimited number of synthetic images of objects of interest that look nearly indistinguishable from real images. Given an example 3D model of an object and an understanding of the satellite sensor, it is possible to generate labeled synthetic datasets for training. *However, to date, efforts to use synthetic images to train CNN models have had limited success.* Neural networks are very efficient at 'learning' what is unique about the objects in the images. In this case, they quickly learn how the synthetic generation process produces the object instances. As a result, the model is unable to recognize the object in *real* images that lack the artifacts of the generation process. The model fails to generalize to the way that the object appears in real images.

As part of a National Geospatial-Intelligence Agency (NGA) Small Business Innovation Research (SBIR) grant, Orbital Insight (OI) and its partner Rendered.ai made significant progress in solving two of the problems by using synthetic images to train CV models: how to modify synthetic images, so the trained model can generalize to real images; and how to use the combination of both a large set of synthetic images and a small set of real examples efficiently to jointly train a model. We demonstrated improved outcomes for object-detection performance through the use of synthetic data.

OI and Rendered.ai identified several techniques that must be used together in order for synthetically generated images to be additive in training detection networks. Specifically, the experiment achieved the best results by combining a rich and robust set of 3D object models with both 2D and 3D simulated backgrounds; variable lighting and coloration; state-of-the-art 'classical' techniques for domain adaptation; and most critically, the use of CycleGAN models to adapt synthetic images to a target imagery domain. In this case, the target domain is the type of images produced by the optical platform used in the Xview[1] dataset. Without this adaptation, the synthetic images were unable to outperform the real images when training the model to detect real-world instances of the objects of interest.

Baseline Model and Dataset

The experiment baseline definition detects a set of 'novel' object classes in a series of satellite images from a small set of labeled images. The experiment uses image data from the popular, public Xview dataset created by the Defense Innovation Unit Experimental and NGA. The experiment aims to detect novel objects in real-world images; specifically, three separate classes of cranes: "crane truck," "mobile crane," and "tower crane." These objects have relatively few training examples in the dataset. In addition, mobile cranes and tower cranes do not look much like any of the basic object types (cars/trucks/planes/ ships) typically studied in synthetic images, so they are good proxies for the novel class problem — which is detecting an object that is not much like anything one has detected before. In contrast, the crane

[1] http://xviewdataset.org/

truck class is a subclass of trucks that look a lot like other types of specialized trucks, such as concrete pump trucks, so utilizing the crane truck serves as a proxy for the novel-subclass problem.

To emulate the problem of finding novel objects from limited examples, this experiment reverses the traditional roles of the dataset split. The smaller Xview validation set serves as this model's training set, while the larger Xview training set is the model's validation set. The smaller validation set represents the scenario of having few examples of the novel classes, while the larger training set represents the set of many general examples. The large set of objects for validation will provide a better understanding of how well the model generalizes to the model classes and supports more efficient evaluation of the different approaches to synthetic imagery generation.

This experiment uses the Faster R-CNN detection model defined in Facebook Research's Detectron2,[2] pre-trained on ImageNet over three epochs. Orbital used transfer learning on the pre-trained model, unfreezing the backbone (Resnet50) up to layer 2. In most cases, the team used default parameters to run a series of experiments that tuned the hyper-parameters on the real imagery for CNN training and then used those settings for the synthetic imagery training and validation. The non-default hyper-parameter settings included:

- Scaling up all images by two times to make use of every pixel in the satellite imagery; because the model's Resnet50 backbone was designed for terrestrial applications, it tends to give up resolution to reduce memory footprint; upscaling the images by two times mitigated this issue and supported training and validation

- Using Learning rate scheduler with Cosine Annealing and warmup to achieve a base learning rate of 0.005

- Modifying the anchor box sizes and aspect ratios as follows to better fit the sizes and shapes of the objects; computing optimal sizes with histograms of object sizes and aspect ratios in the training set

  SIZES: [[25], [50], [100], [200], [400]]
  ASPECT_RATIOS: [[0.33], [0.5], [1.0], [2.0], [3.0]]

- Augmenting training data, such as brightness changes (0.8-1.2 range), contrast changes (0.8-1.2), saturation (0.8-1.2), random horizontal and vertical flips, and rotations of plus/minus 20 degrees; rotations are particularly important to avoid over-fitting the small training sets

Synthetic Data Set Creation

The experiment leveraged commercial capabilities for the synthetic dataset tools provided by Rendered.ai. The Rendered.ai platform provides a simple way to generate data with different characteristics in a reproducible way. The platform makes it easy to configure datasets with different properties, scenarios, and effects efficiently, which makes rapid iteration and experimentation possible. Configurations are encoded and stored, allowing for reproducibility, which is critical for machine-learning research and development. Orbital performed extensive experimentation to produce over 500,000 synthetic images. Experimentation supported both improved outcomes and a deeper understanding of what did and didn't work. The team discovered that the use of Domain Transfer techniques on the synthetic imagery is a critical step and explored several approaches to Domain Transfer. The GAN-based algorithms performed the best.

The datasets and the workflow tools that produced these results are available upon request to the customer in a collaborative cloud environment that also supports new dataset creation and experimentation.

[2] https://ai.facebook.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/

Figure 1: Commercially available, cloud-native synthetic data software was used to achieve improvements in average precision scores for important objects of interest. Workflow summary from left to right: dataset generation configuration, training dataset, trained model. We point the way forward for further advances in these capabilities.

Orbital performed a number of experiments to optimize various parameters with significant impact on model training and prediction accuracy, as briefly summarized here:

- **Color space.** Synthetic data often do not match the color of real objects; tuning is required to mimic the color distribution of real data. The team experimented with adjusting and removing the hue of objects in the synthetic data. The results show color is not a feature that plays an important role for detection.
- **Real vs synthetic background.** Synthetic data generation can be a resource- and compute-intensive effort. With the abundance of real Earth-observation imagery available, this experiment aimed to understand how effectively real imagery can serve as background for synthetic objects of interest. The team leveraged existing xView data with no objects of interest as background and tested this against fully synthetic background for the various scenarios in the xView dataset, such as urban, suburban, and mining. Without domain adaptation, both datasets performed significantly worse. However, with domain adaptation, synthetic data generated with 3D background performed better than 2D background for the tower crane class.
- **Distractor objects.** The team hypothesized that the model would overfit on inherent artifacts in synthetic data generated using an xView background. To test this, the experiment added distractor synthetic objects similar in the size of the objects of interest to the generated images. The results showed that the distractor objects did not make a noticeable change in model performance.
- **Distribution of objects of interests.** Dealing with imbalanced datasets is common in machine learning. The benefit of synthetic data is better control over the generation of balanced and imbalanced datasets. The team created a dataset that randomized the distribution of the three classes of objects of interest as well as one with a uniform distribution.

Image Domain Adaptation

For the CNN model to learn to identify objects in real images, the synthetic training data in feature space must come from the same statistical distribution—or domain— as the real imagery. CNN models can quickly learn a dataset's distribution, and if the synthetic differs from real imagery, the model will not be able to detect objects in real images. Therefore, the domain of the synthetic images must be adapted to match that of the real images.

Orbital focused on two approaches to domain adaptation to improve real data performance for models trained on synthetic data: traditional and neural-based domain adaptation techniques. For the traditional approach, the team applied adjustments and corrections in the Fourier space and color space and used various filters and blending techniques. The team used CycleGAN image-to-image translation to adapt synthetic images to improve the ability of the synthetic training dataset to generalize to the real-world test

dataset. The CycleGAN model was created to translate synthetic images to match the Xview real-world images. The CycleGAN model was trained to match the Xview corpus of images. The experiment determined the optimal number of cycles is eight epochs.

Examples of Synthetic Images

The following illustrations show a few examples of the synthetic images created for the experiment. While the placement of multiple synthetic objects in these environments looks wrong to humans, they actually help the model learn to recognize the objects in many different scenarios. Despite the use of many traditional domain- adaptation techniques to make the images look more realistic, they still appear as synthetic to humans—and to the model.

Examples of Domain-Adapted Synthetic Images

These examples of domain-adapted versions of the synthetic images were created using the CycleGAN model trained on the Xview dataset. The images now appear more realistic to both humans and the neural network. Two of the four images are synthetic, and two are real.

Results

Baseline Line Model

Figure 2 compares experiments with synthetic images to the baseline model.

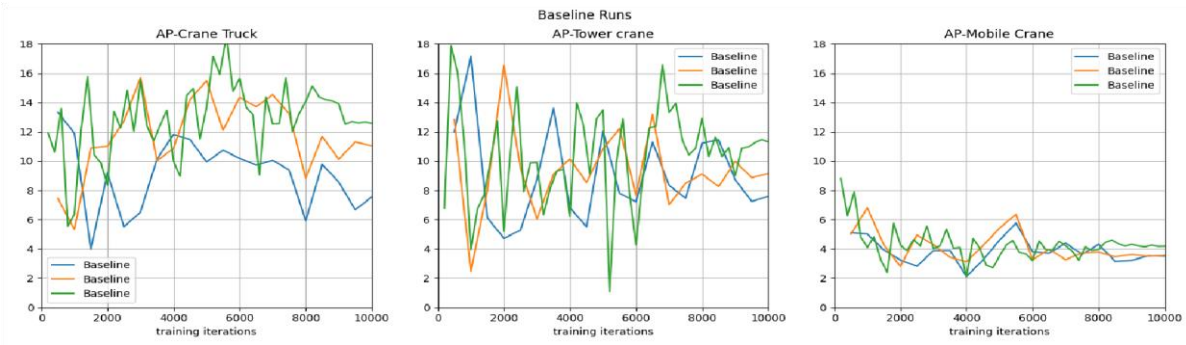


Figure 2. Baseline performance achieved by training on only the small amount of real data available was very low across all three categories; most notably, mobile cranes. The variance in results is due to stochastic training.

Orbital created a series of synthetic datasets, each with different traditional and deep learning-based domain adaptation approaches. Models trained with synthetic images improved based on the adaptation methods used. Incremental training the model with real images after training with CycleGAN-adapted synthetic images produced better results than the baseline.

The average precision (AP) scores are calculated with an average of bounding box Intersection-OverUnion (IOU) over a range of 0.15–0.50 instead of the traditional range of 0.5–0.95. Figure 2 shows the AP scores for the three object classes as a function of training iteration over three separate training runs. This provides the baseline for the AP scores and their variance from run to run due to the statistical nature of stochastic gradient descent. The illustrations show the mobile crane class has relatively poor AP. The perrun variance for each class is several percentage points.

Domain Adaptation

Training on synthetic data alone does not generalize to real examples and performs worse than training on the small real image set, even with thousands of unique synthetic examples. The CycleGAN versions of the synthetic datasets improve performance over original synthetic images, in some cases (the tower crane class) surpassing the baseline on real training images. The crane truck class AP was very low for Synth0 and Synth1 datasets, even after CycleGAN adaptation. However, when the models were improved in Synth3, the CycleGAN versions of the class approached the level of the real images, as shown in Figure 3.

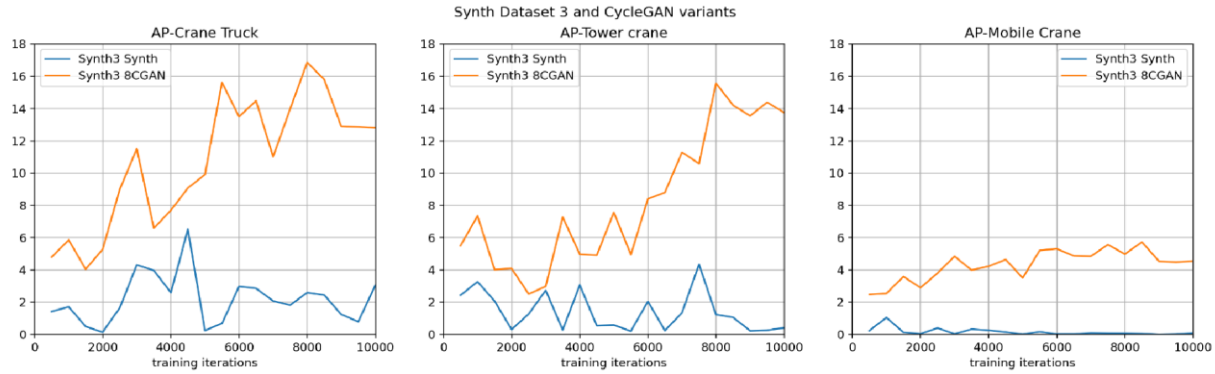


Figure 3. AP of the three target classes on a synthetic dataset (Synth3) and the CycleGAN domain-adapted version of the same dataset; showing the improvement in the ability of the model to learn when using CycleGAN-adapted images for training.

Incremental Training

The team experimented with different ways to train a model using both the large domain-adapted synthetic image dataset and the much smaller real examples. The models trained on the combined dataset did about as well as the average of the datasets individually, sometimes actually decreasing the model's AP scores compared to training on either dataset alone.

Training the model on the synthetic dataset first and then incrementally training on the real training data improved scores above either model and even above the real data baseline. This created a solution that could use synthetic training data and get results better than real data.

The results of incremental training is shown in Figure 4. In this example, the model trained for 14,000 iterations using the combined 8-cycleGAN dataset, then continued training for 10,000 iterations using only the real training data. For each class, the incremental training improved the AP over the CycleGAN training and surpassed the baseline real APs.

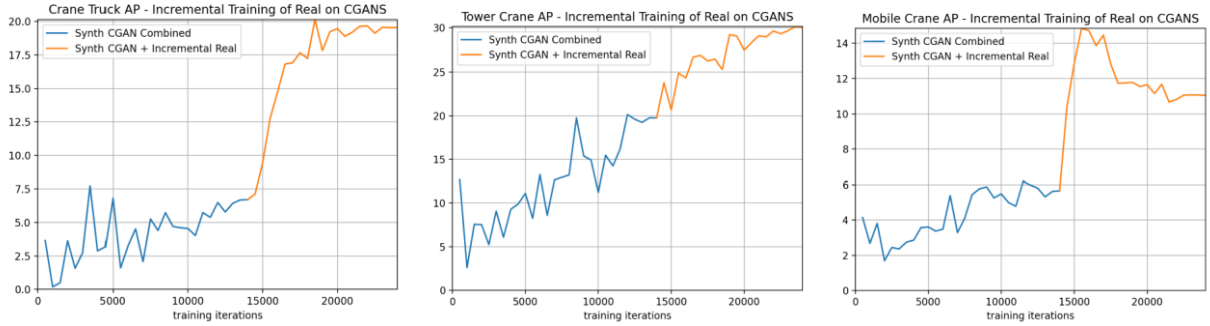
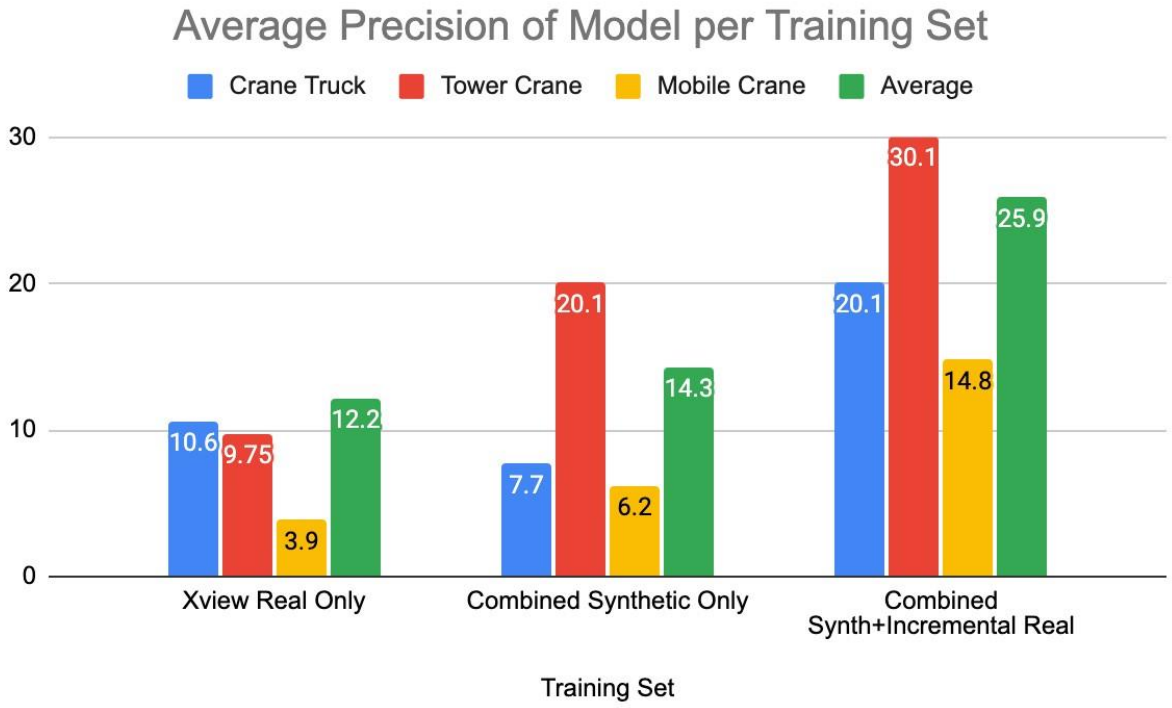


Figure 4. AP of the three classes after training on the combination of all of the CycleGAN synthetic datasets and then incrementally training on the real data; showing improvement across all three classes and exceeding the levels of the baseline results on real data.

The following results charts and table (Table 1) summarize the AP scores for each type of training set: real only, synthetic only, and combined real and synthetic with incremental training.



Training Set	Crane Truck AP	Tower Crane AP	Mobile Crane AP	AP15
Xview Real Only	10.6	9.75	3.9	12.2
Combined Synthetic Only	7.7	20.1	6.2	14.3
Combined Synthetic + Incremental Training on Real	**20.1**	**30.1**	**14.8**	**25.9**

Table 1: Maximum AP scores for models trained on real-only, combination of three 8-cycleGAN datasets, and using incremental real after combined synthetic training:

Conclusion and Next Steps

Orbital concludes it is technically feasible to create synthetic images for training detection models that, when combined with a small set of real training images, can train a model to identify the classes in the training set. With further experimentation, it should be possible to improve the accuracy of the model through a better understanding of how the CycleGAN domain adaptation is operating on the synthetic images and how the combination of the small real dataset improves upon the synthetics during the second phase of the training.
The results show improvements on both novel classes and on novel sub-classes. The term novel classes indicates object types that do not look like a previously studied class (e.g., tower cranes that have a unique shape). Novel classes are challenging in that some may require that the detector learn new features in the backbone (e.g., the long truss structures of tower cranes). Novel subclasses are classes such as crane trucks, which are trucks and a well-studied class and possess particular and distinct features (in this case, a large extensible boom). The challenge in these cases is to accurately distinguish the novel subclass from all of the other, often much more common, kinds of trucks.

While AP scores clearly improved in these experiments, both over pre-existing work and realimagery-only baselines, they are still below what would be required for a deployed system. Orbital recommends the following next steps to improve detection scores to a level required for a deployed system:

- Higher quality and quantity for real training and validation imagery
- Higher quality and quantity for object models for the simulation engine
- More and better image modifiers and classical domain-adaptation techniques
- Direct research on CycleGans for creating simulated imagery
- Additional research on neural network architectures for training on simulated and mixed-simulated/real imagery

The overarching goal of all the tasks is to demonstrate how to build detectors that are sufficiently accurate to be used in practice; and to benchmark how much real training imagery is required to reach this level of accuracy.

The end game of the work is to create a solution that takes in object models—essentially, CAD drawings—for objects of interest and produces detectors for those objects that can be applied to satellite and aerial imagery. This will require considerable additional work but, once complete, would support the quick development and deployment of detectors for many kinds of objects that cannot be accurately identified by the AI systems of today.