Closing the Domain Gap – Blended Synthetic Imagery for Climate Object Detection

Caleb Kornfein, Frank Willard[*], Caroline Tang*, Yuxi Long*,
Saksham Jain*, Jordan Malof*, Simiao Ren*, Kyle Bradbury*
Duke University, Durham, NC 27701, USA
{caleb.kornfein, frankie.willard, caroline.tang, yuxi.long,
saksham.jain, jordan.malof, simiao.ren, kyle.bradbury}@duke.edu

Abstract

Object detection models have great potential to increase both the frequency and cost-efficiency of assessing climate-relevant infrastructure in satellite imagery. However, model performance can suffer when models are applied to stylistically different geographies. We propose a technique to generate synthetic imagery using minimal labeled examples of the target object at a low computational cost. Our technique blends example objects onto unlabeled images of the target domain. We show that including these synthetic images improves the average precision of a YOLOv3 object detection model when compared to a baseline and other popular domain adaptation techniques.

1 Introduction

From power plants to wildfires, many of the causes and consequences of climate change are visible from above. Accurate geospatial information about the causes of climate change including energy infrastructure systems is critical to planning climate change mitigation and adaptation strategies. However, spatial data on current energy infrastructure is often lacking – the data may not be publicly available, may be incomplete, or may not contain sufficient granularity [1]. Recent research has demonstrated the potential of using satellite imagery to fill the data gaps by monitoring energy systems at unprecedented frequencies and scale [2] [3]. Two remaining challenges to applying these techniques at scale include: (1) a lack of large datasets with labeled data for relevant applications, and (2) the challenge of applying these techniques across diverse geographic domains.

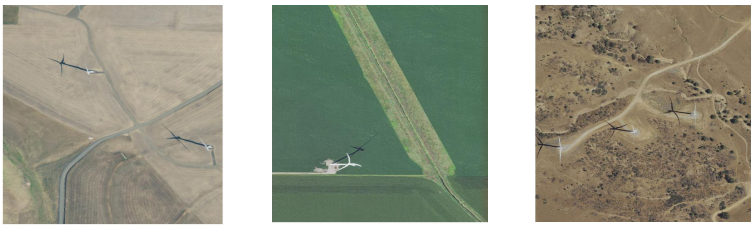

Figure 1: Overhead images of wind turbines from different regions of the United States

The first issue is the high number of labeled examples required to successfully train a model. Building the right dataset can be challenging because of the labor-intensive nature of accurate data labeling

[*]Equal contribution.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

[4]. This is of particular significance in the case where the object of interest is rare and few labeled instances may exist, which is arguably often the case for data related to energy and climate systems.

The second challenge is to apply these techniques across geographic domains. While we are often limited to acquiring image data from a narrow set of geographies, we may want to apply models more broadly to different geographic regions. This difference in the visual properties of the images we use to train versus those we use to test our models is known as the domain gap and deep neural network models often perform unreliably in such conditions [5]. Our experiments trained object detection models to detect one type of energy infrastructure, wind turbines, and differentiated between *within-domain* experiments in which we create a wind turbine detection model using training and testing imagery from the *same* geographic domain, versus *cross-domain* experiments where we train our model on one domain and test it on a *different* domain. As an example, a within-domain experiment may train on images of wind turbines from the Northwest U.S. and validate on other images from the Northwest. In contrast, a cross-domain experiment may train on images from the Northwest and test on images from the Southwest. We expect object detection performance to suffer in cross-domain contexts due to the presence of a domain gap. We experimentally demonstrated the existence of a domain gap in our dataset in Figure 2 by showing how, on average, within-domain performance of our object detection models was far greater than cross-domain performance for each test domain.

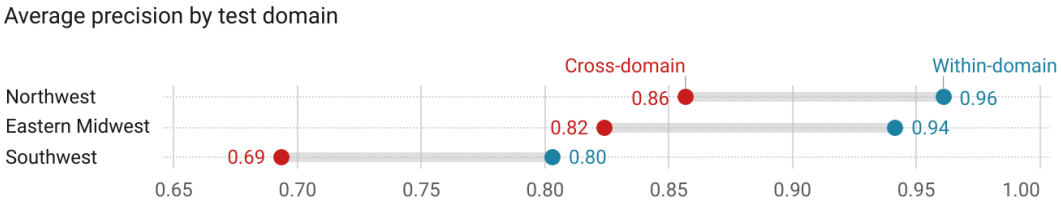

Figure 2: Baseline experimental results demonstrating evidence of the domain gap. The gap in performance between within-domain and cross-domain settings is shown in this figure as the distance between the red and blue points (and further described in section 3)

Two approaches to domain adaptation in remote sensing imagery are transforming the images in a dataset to lessen the differences between the training and testing domains or supplementing the dataset with synthetic imagery that helps to close that gap (e.g., Synthinel-1 [6], SIMPL [7], [8]). Color based transformations leave image content unchanged but vary the color of pixels in an image, mapping pixels of each color to a new value. For example, histogram equalization adjusts image pixel intensity values such that they follow a uniform distribution to standardize image appearance regardless of domain. In contrast, generative methods create synthetic images that may modify both the pixel values and the information content of an image. Examples of generative methods that have been applied for domain adaptation include CyCADA [9] and CycleGAN [10]. We compare our approach to a selection of these methods.

We propose a domain adaptation method based on generating synthetic overhead images designed to require few labeled examples. Our technique takes advantage of unlabeled overhead images to reduce the number of labeled examples required. Unlabeled images are often easily acquirable through public datasets such as the National Agriculture Imagery Program (NAIP) [11]. Our experiments show the potential of our technique to improve downstream model performance as compared to other domain adaptation techniques in this space – especially in situations with limited training data and when applying a model to new geographies.

2 Methodology

2.1 Experimental dataset creation

To test our domain adaptation technique, we created a dataset of overhead images containing wind turbines from three distinct geographic domains. Wind turbines were selected as an example of energy infrastructure because they are relatively homogeneous in appearance (minimizing intra-class variance), found in a diverse variety of geographies and contexts (mountains, fields, etc.), and are still relatively rare geospatially.

We collected images from the Northwest, Southwest, and Eastern Midwest regions of the United States. These regions were chosen with the intention of creating visually distinct domains, as shown in Figure 1. Wind turbine coordinates were sampled from the U.S. Wind Turbine Database [12]. Overhead images were collected from the National Agriculture Imagery Program dataset (NAIP) [11] using Google Earth Engine [13]. Each image was 608x608 pixels with a resolution of 1m/pixel. For each domain we also included unlabeled (background) images, which are visually similar to the labeled examples but do not contain wind turbines. The final dataset contained 988 labeled images with turbines and 1,015 background images. See Appendix A for further details on data collection.

2.2 Generating synthetic blended images

Our image generation process aimed to produce synthetic images that were similar to real labeled data from the target domain. We utilized the GP-GAN image blending model [14] to blend unlabeled target domain data together with source domain target objects. This process consisted of four steps, as shown in Figure 3: (1) sample a random background image (without wind turbines) from the *target* domain, (2) sample turbines from the *source* domain, (3) randomize the location, orientation, and size of the objects on a blank canvas, and (4) blend the background image and the source domain turbines together using GP-GAN.

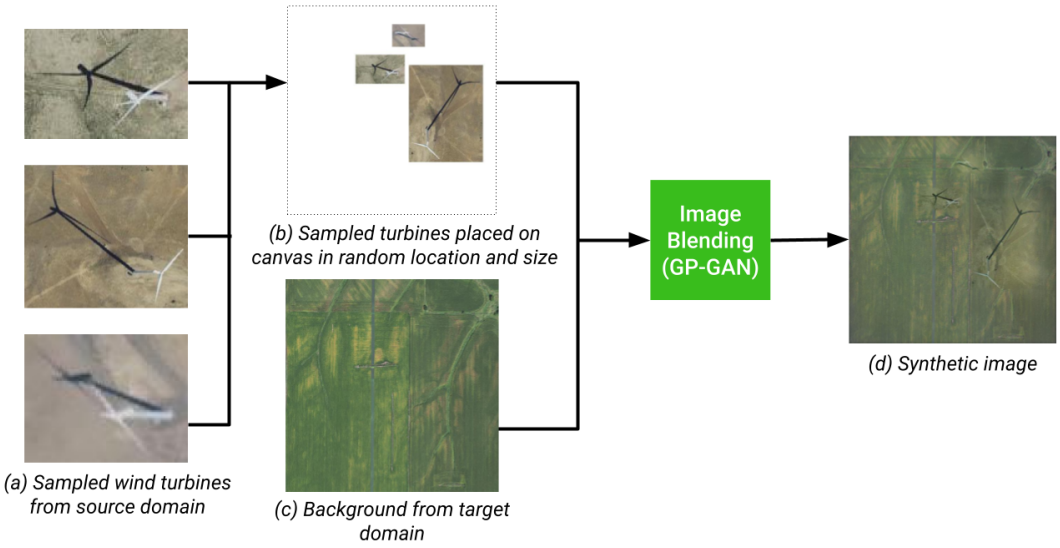

Figure 3: Synthetic image generation process

This process is customizable with hyperparameters such as the object density, spacing, and size. For our experiments, we ensured that the wind turbines never overlapped and that the object density was set to the 90th percentile of object density from the source domain.

2.3 Object detection model and evaluation metrics

We used YOLOv3 [15] as our wind turbine object detection model. With the model, we used mixed batch training to ensure a fixed 7-to-1 ratio of baseline-to-supplementary data in each batch, increasing our control over the influence of the supplemental data within each experiment. We evaluated model performance using average precision (AP).

3 Results and Discussion

Our experiments simulated detecting a rare energy object: we assumed minimal access to labeled images from the source domain and no labels from the target domain. The experiments compared the various domain adaptation techniques by investigating whether the addition of different types of supplementary imagery (including our synthetic images) could improve wind turbine detection performance across domains.

The baseline experiment was trained on 100 labeled source domain images and tested on 100 labeled target domain images. Each domain adaptation technique was tested by training on 100 labeled source domain images *supplemented* with an additional 100 images from a domain adaptation technique and tested on the same 100 labeled target domain images (see Appendix B, Figure 6). We ran this for our synthetic image blending technique, introduced here, and compared it to other domain adaptation techniques including Histogram Equalization, Histogram Matching, GrayWorld, CyCADA, and CycleGAN. Each experiment tested all of the 9 possible source/target domain pairings across the three domains: Northwest, Eastern Midwest, and Southwest. For each pair of domains, 5 YOLOv3 models were trained to estimate model variance. All models were trained for 300 epochs with a batch size of 8. The results of these experiments are shown in Figure 4.

We added to the above baseline and domain adaptation techniques two additional experimental points of comparison that were modifications of the baseline experiment. First, to estimate an upper bound on performance to training with 100 supplemental images, we supplemented the baseline training process with 100 additional real images from the *target* domain. Second, to explore if simply adding target domain imagery improved performance, regardless of the presence of wind turbines in the imagery, we supplemented the baseline experiment with 100 target domain background images (unlabeled target domain images without wind turbines present used for image blending). The results of these two experiments are shown in Figure 4 are shown in light grey.

These experiments collectively tested whether adding each set of supplemental imagery improved cross-domain model performance when faced with highly-limited training data availability. The results indicate that adding synthetically blended imagery is able to produce the greatest improvement in average precision of the techniques compared in this study. On average, our synthetic image blending technique outperformed baseline trials by 8.23% in cross-domain pairings and 2.21% in within-domain pairings. Additional exploration of these results can be found in Appendix C and D.

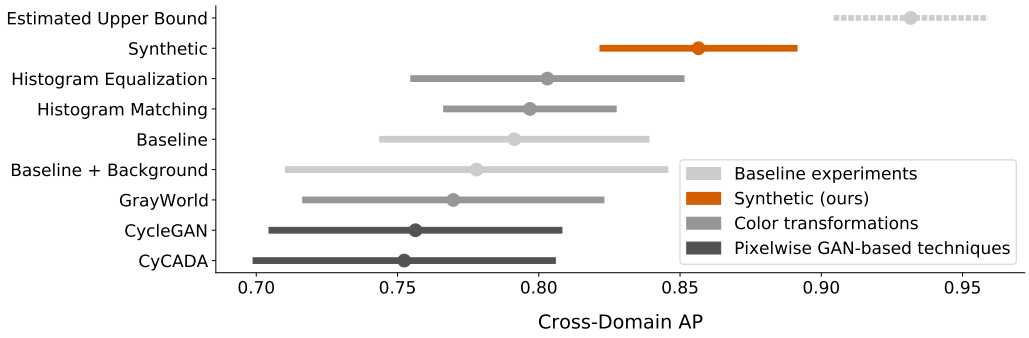

Figure 4: Experimental results comparing average cross-domain AP $\pm$ 2 std. dev.

4 Conclusions

The mapping of climate relevant infrastructure can improve the assessment of key resources, but global applications have been impeded by the challenges of geographic domain adaptation and a lack of training data. Our experiments provide evidence that supplementing training data with synthetically blended imagery can improve domain adaptation while requiring minimal time, no proprietary software, and few object examples. This may aid in scaling up mapping to larger applications. Future work could further refine our synthetic imagery quality by adjusting the density or size of objects blended or by validating our method on different energy infrastructures or domains.

Acknowledgments and Disclosure of Funding

We would like to acknowledge the support of the Duke University Data+ and Bass Connections programs as well as the Nicholas Institute of Energy, Environment and Sustainability and the Rhodes Information Initiative . We thank our collaborators Wei Hu, Madeleine Jones, Alena Zhang, Maddie Rubin, Alexander Kumar, Aya Lahlou, Boya (Jennie) Sun, and Katie Wu.

References

[1] Dan Stowell, Jack Kelly, Damien Tanner, Jamie Taylor, Ethan Jones, James Geddes, and Ed Chalstrey. A harmonised, high-coverage, open dataset of solar photovoltaic installations in the uk. *Scientific Data*, 7(1):394, 2020.

[2] Simiao Ren, Wayne Hu, Kyle Bradbury, Dylan Harrison-Atlas, Laura Malaguzzi Valeri, Brian Murray, and Jordan M. Malof. Automated extraction of energy systems information from remotely sensed data: A review and analysis. *Applied Energy*, 326:119876, 2022.

[3] Priya L. Donti and J. Zico Kolter. Machine learning for sustainable energy systems. *Annual Review of Environment and Resources*, 46(1):719–747, 2021.

[4] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International Conference on Artificial Neural Networks*, 2018.

[5] Devis Tuia, Claudio Persello, and Lorenzo Bruzzone. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, 2016.

[6] Fanjie Kong, Bohao Huang, Kyle Bradbury, and Jordan Malof. The synthinel-1 dataset: A collection of high resolution synthetic overhead imagery for building segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1814–1823, 2020.

[7] Yang Xu, Bohao Huang, Xiong Luo, Kyle Bradbury, and Jordan M. Malof. Simpl: Generating synthetic overhead imagery to address custom zero-shot and few-shot detection problems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:4386–4396, 2022.

[8] Wei Hu, Tyler Feldman, Eddy Lin, Jose Luis Moscoso, Yanchen J Ou, Natalie Tarn, Baoyan Ye, Wendy Zhang, Jordan Malof, and Kyle Bradbury. Synthetic imagery aided geographic domain adaptation for rare energy infrastructure detection in remotely sensed imagery. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.

[9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[10] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[11] USDA Farm Production and Geospatial Enterprise Operations Conservation Business Center. National agriculture imagery program.

[12] B Hoen, J Diffendorfer, J Rand, L Kramer, C Garrity, and H Hunt. United states wind turbine database (version 3.3, january 2021). *US Geological Survey, American Wind Energy Association, and Lawrence Berkeley National Laboratory Data Release*, 2018.

[13] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017.

[14] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Gp-gan: Towards realistic high-resolution image blending. *ACMMM*, 2019.

[15] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

A Dataset creation

Our wind turbine imagery dataset consists of overhead imagery from the Northwest, Southwest, and Eastern Midwest domains of the United States shown above in Figure 5. Images were created for each domain by:

1. Clustering the U.S. Wind Turbine Database coordinates of each region via DBSCAN
2. Selecting subsets of candidate test coordinates and training coordinates using stratified random sampling to ensure representative sampling within each region
3. Capturing 608m x 608m overhead images using Google Earth Engine's publicly available NAIP dataset
4. Shifting the coordinate center of each image by 75 meters to ensure that a turbine was not directly in the center of the image each time
5. Each image was quality checked to ensure that wind turbines were present in the image and that the training and test datasets were mutually exclusive

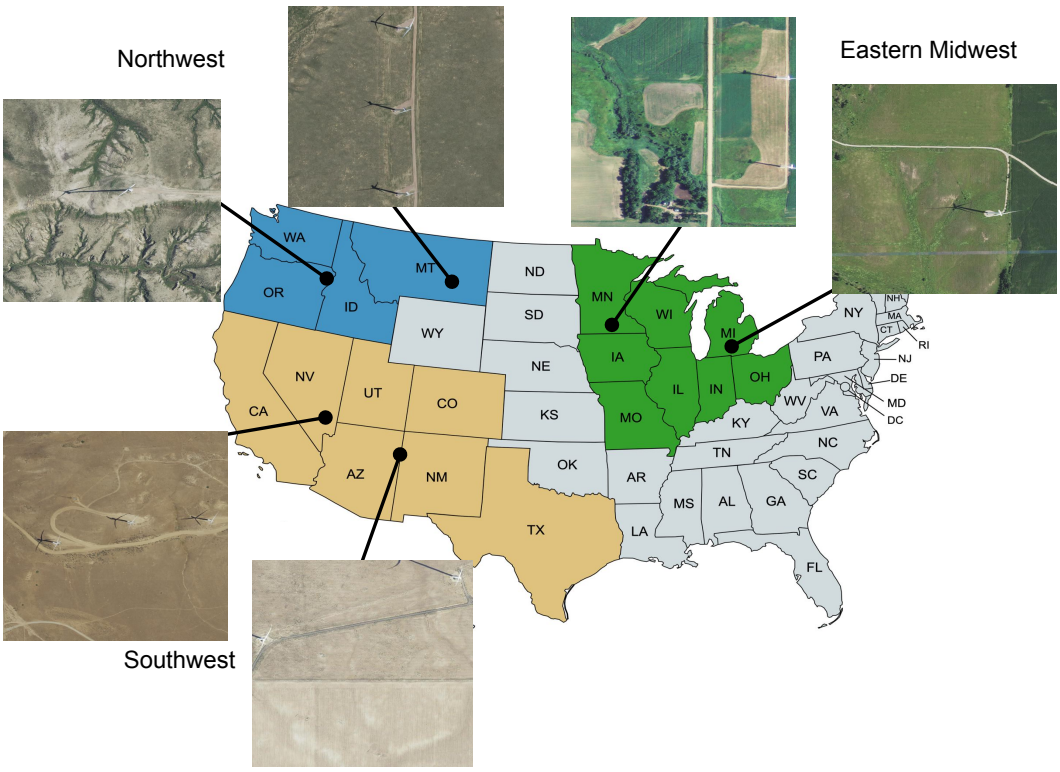

Figure 5: Sample images and corresponding locations from the Northwest, Southwest, and Eastern Midwest regions of the wind turbine dataset

To collect the background images without wind turbines (to which wind turbines could be added through blending with GP-GAN), overhead images were captured at a small distance between 4km and 6km away from a known turbine location. These images were manually inspected to ensure that no wind turbines were present. The close distance ensured visual similarity to the other target domain images.

B Experimental setup

Code for recreating our experiments and results can be found at: https://github.com/energydatalab/closing-the-domain-gap. Our dataset can be downloaded at: https://zenodo.org/record/7385227.Y4hf–zMKw5.

Figure 6 shows our general experimental setup: (1) a baseline without 100 source domain images and no supplemental images tested on a dataset of target domain images and (2) each experimental condition with the same 100 source domain images supplemented with images from each experimental condition, tested on the same target domain images.

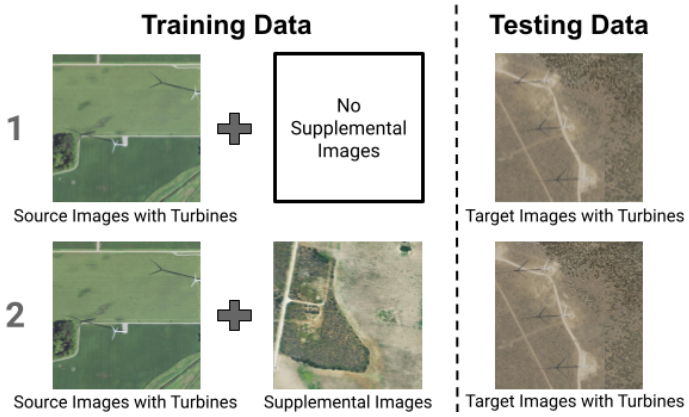

Figure 6: Experimental setup types

For the experiments which contained a supplement to the training set (type 2), we ensured the Yolov3 models trained with a mixed batch 7-to-1 ratio of train and supplement image in each minibatch. This ratio was chosen as the lowest possible ratio in a minibatch size of 8 that would still allow the consistent influence of the supplemental imagery. Each epoch consisted of a pass through 200 images at the given train-to-supplement ratio.

C Baseline and synthetic results

The full baseline results in Table 1 highlight the domain gap present in the dataset. Overall, baseline within-domain trials achieved an average AP of 0.901 while baseline cross-domain trials achieved an average AP of 0.791. For each domain pair, the addition of synthetic images improved AP. This was especially true in a cross-domain context: on average, synthetic cross-domain trials achieved a 0.065 higher AP than baseline cross-domain trials, while synthetic within-domain trials achieved a 0.020 higher AP than baseline within-domain trials.

Table 1: Baseline and summary synthetic experiment results compared via average trial AP.

Source Domain	Target Domain	Baseline $\pm$ 2 SD	Synthetic $\pm$ 2 SD
EM		0.941 $\pm$0.029	0.964 $\pm$0.024
NW	EM	0.894 $\pm$0.025	0.936 $\pm$0.022
SW		0.647 $\pm$0.055	0.800 $\pm$0.026
EM		0.905 $\pm$0.058	0.930 $\pm$0.037
NW	NW	0.960 $\pm$0.019	0.963 $\pm$0.007
SW		0.739 $\pm$0.050	0.772 $\pm$0.032
EM		0.742 $\pm$0.067	0.836 $\pm$0.049
NW	SW	0.818 $\pm$0.028	0.862 $\pm$0.041
SW		0.802 $\pm$0.043	0.837 $\pm$0.030
Within-Domain Average		0.901 $\pm$0.031	0.921 $\pm$0.020
Cross-Domain Average		0.791 $\pm$0.047	0.856 $\pm$0.035

D Full experiment results

In addition to AP, we measured experiment results using the percentage Closure of Domain Gap (CDG) [8]. CDG measures the percent closure of the gap between the within-domain and cross-domain trials that the current experiment recovers:

$$CDG = \frac{AP(\text{cross-domain experiment}) - AP(\text{cross-domain baseline})}{AP(\text{within-domain baseline}) - AP(\text{cross-domain baseline})} \quad (1)$$

A full comparison of techniques using the CDG metric as well as both cross-domain and within-domain AP are shown in Table 2. Our synthetic imagery achieved the highest cross-domain and within-domain AP out of any experiment except the estimated upper bound experiment. Additionally, our synthetic imagery achieved the highest closure of the domain gap out of any technique compared in this study.

Table 2: Comparison of synthetic imagery to alternative domain adaptation techniques. Outcomes measured in average trial AP and split between the cross-domain and within-domain trials.

Experiment	Cross-domain	Within-domain	CDG
Synthetic	**0.856**	**0.921**	**60.2%**
CyCADA	0.752	0.877	-39.2%
CycleGAN	0.756	0.889	-30.7%
GrayWorld	0.769	0.891	-15.1%
Histogram Matching	0.796	0.906	13.1%
Histogram Equalization	0.803	0.909	16.7%
Estimated Upper Bound	0.931	0.938	139.4%
Baseline	0.791	0.901	0.0%
Baseline + Background	0.778	0.903	-16.3%

E Domain adaptation technique examples

Figure 7 shows examples of each of the domain adaptation techniques for generating supplemental training images that were included in this study. The top row shows the original real images upon which each of the subsequent images was based to give a qualitative comparison of each technique.

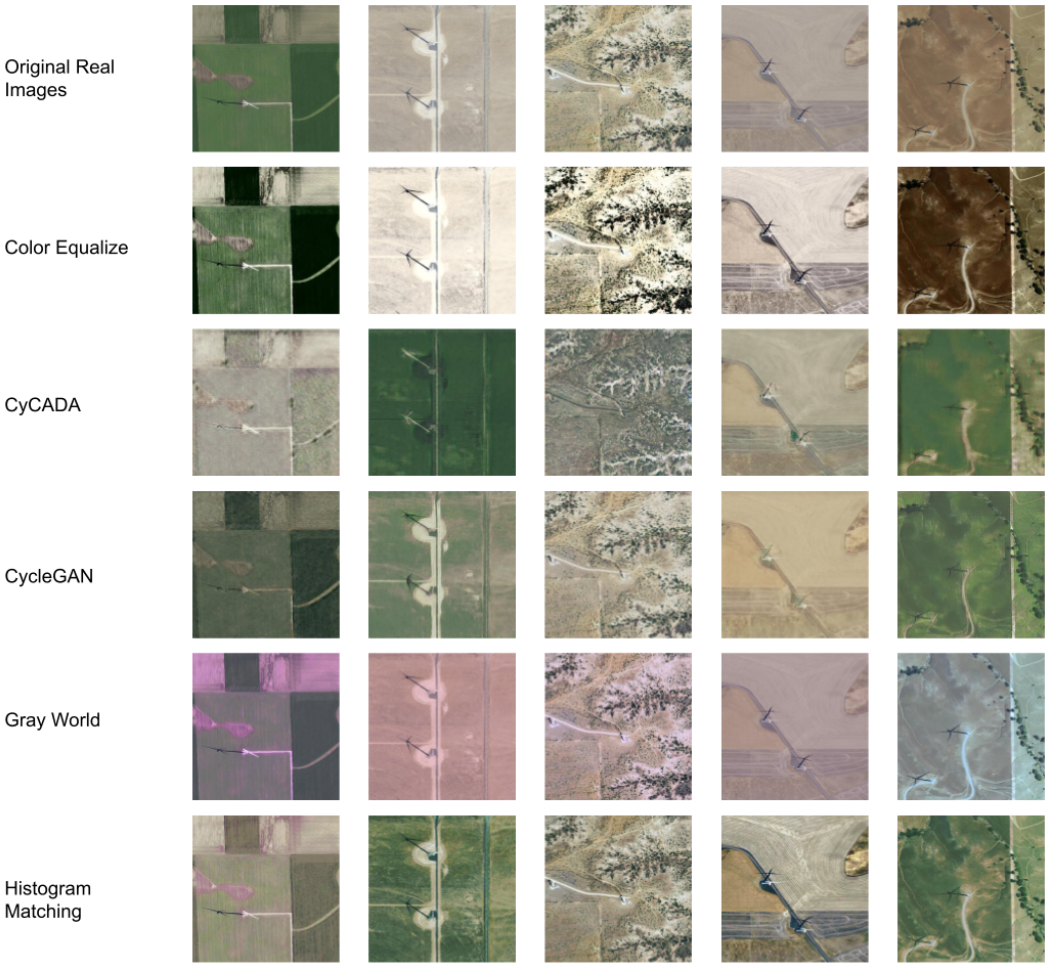

Figure 7: Examples of each domain adaptation technique

F Background image and GP-GAN synthetic image examples

Figure 8 shows examples of images to which wind turbines were added through GP-GAN blending.

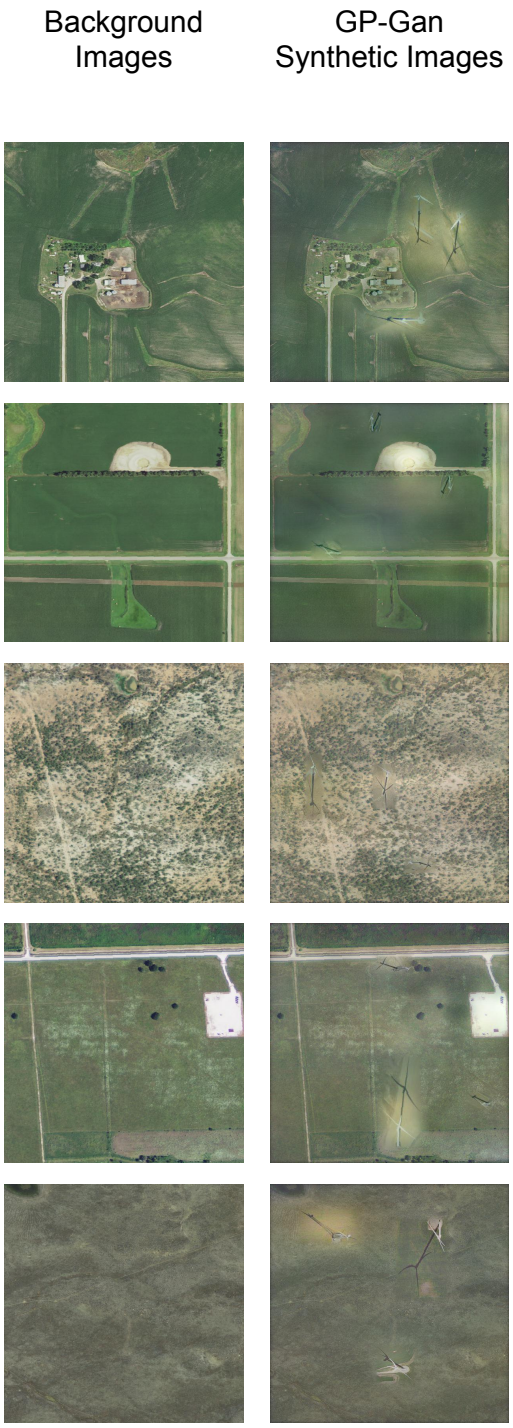

Figure 8: Examples of background overhead images (left) and the blended synthetic images with added wind turbines (right)