Wind Turbine Detection with Synthetic Overhead Imagery

Wei Hu[1], Tyler Feldman[2], Yanchen J. Ou[3,4], Natalie Tarn[3], Baoyan Ye[5], Yang Xu[2,6], Jordan M. Malof[2], Kyle Bradbury[1,2]

[1]Energy Initiative, [2]Department of Electrical and Computer Engineering, [3]Department of Computer Science, [4]Department of Statistical Science, [5]Trinity College
Duke University, Durham, NC 27701, USA

[6]School of Computer and Communication Engineering
University of Science and Technology Beijing, Beijing 100083, China

ABSTRACT

Automatic object detection in overhead imagery is greatly increasing the pace at which we learn about anthropic activity across diverse fields such as economics, environmental management, and engineering. Properly-trained object detection models save significant amounts of human labor when it comes to finding objects, especially rare objects, in overhead imagery. However, applying such techniques to find rare objects typically requires a large amount of labeled imagery data (typically requiring expensive manual labeling). We generate synthetic imagery to reduce the amount of manually labeled imagery required to train models, particularly for data-constrained applications. This approach takes real, unlabeled overhead imagery and inserts artificial 3D models of objects onto the imagery. To evaluate this technique, we collected a baseline dataset of overhead imagery with wind turbines that have been manually labeled and overhead imagery that does not contain wind turbines. We then add synthetic imagery to some of the unlabeled data to create a synthetic dataset. Our results indicate that adding synthetic imagery in training achieving higher levels of recall for similar levels of precision, outperforming the baseline of only real imagery.

Index Terms— *Computer vision, object detection, synthetic imagery, electricity infrastructure, remote sensing, wind turbine*

1. INTRODUCTION

Automatic object detection methods in overhead imagery are being deployed for applications across many fields to save time and human effort in better understanding our society and ecosystem. In recent years, convolutional neural networks (CNNs) based object detection algorithms have been extensively used on overhead imagery because its capability of processing multi-band image data. Very high resolution remote sensing imagery (spatial resolution higher than 1 m/pixel) enables object detection from overhead imagery to extract information about a diverse range of small objects that are usually not identifiable from most satellite imagery, including trees, ground vehicles, etc. [1]. However, training a successful CNN-based object detection model requires a large amount of annotated imagery as training samples, which normally requires manual labeling; a process which is time consuming and expensive.

Techniques like transfer learning can reduce the amount of imagery that needs to be manually labeled, but some annotated imagery is still required [2]. A recent paper proposed an approach where synthetic imagery is used to augment real-world training imagery [3]. The synthetic imagery is created by adding artificial 3D models of objects of interest to real background imagery by 3D rendering. It can inexpensively create training imagery with great variety and since both the environment and the objects in it are synthetically created and because of that, all target object locations are known. This means that all synthetic imagery is automatically labeled and does not require human labor for labeling.

In this paper, we extend this idea to explore whether adding synthetic objects to real background imagery will improve a model's object detection performance in training data constrained settings. The benefit of this approach is that the resulting imagery is more photorealistic. Furthermore, by adding synthetic objects to unlabeled imagery from the target environment, we obtain synthetic imagery that better resembles the testing imagery [3]. We frame this problem around wind turbine identification to demonstrate the approach, which are rare objects in overhead imagery and are good proxy examples for other visible objects that are rare, but yet prominently visible in overhead imagery. We collect a dataset that consists of diverse real-world imagery and contains images from areas known to contain wind turbines, having retrieved those locations from the United States Environmental Protection Agency's eGRID database [4].

We curated two datasets for our experiments in this paper: (1) a baseline dataset of real overhead imagery that contains images that includes the object of interest (wind turbines); and a synthetically-augmented dataset where 3D models of wind turbines are artificially inserted into real background images that do not contain real wind turbines. We then compared the performance of an object detection algorithm trained on each of these datasets.

2. METHODS

In this section we describe the preparation of our two experimental datasets, and our wind turbine detection model.

2.1. Baseline real-world imagery dataset preparation

We created the baseline dataset by sampling from the Power Plant Satellite Imagery Dataset [5] which consists of 1 m/pixel aerial imagery from the National Agriculture Imagery Program (NAIP) dataset [6] and annotations for wind turbines in those imagery. A total of 1,750 images (608 × 608 in pixels each) from 33 US states were randomly selected in our baseline dataset and the 15 most sampled states are shown in Figure 1. We then split all images to create a training set (1,337 images, 80% of the data) and a testing set (413 images, 20% of the data).

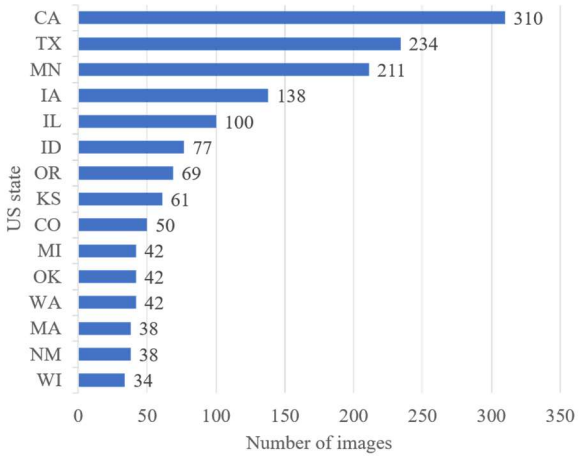

Figure 1. Number of images in the baseline dataset by US state.

2.2. Synthetic imagery dataset preparation

We then use the ArcGIS CityEngine software to create synthetic imagery by augmenting the baseline dataset with artificial images of wind turbine rendered from a 3D model [7]. To create a synthetic image, we first took a new image from the NAIP dataset as the background and then use CityEngine to place 3D wind turbine models on top of the background image. After the 3D wind turbines were added to the background image, we took an overhead photo of the generated 3D scene to create a 2D synthetic image that

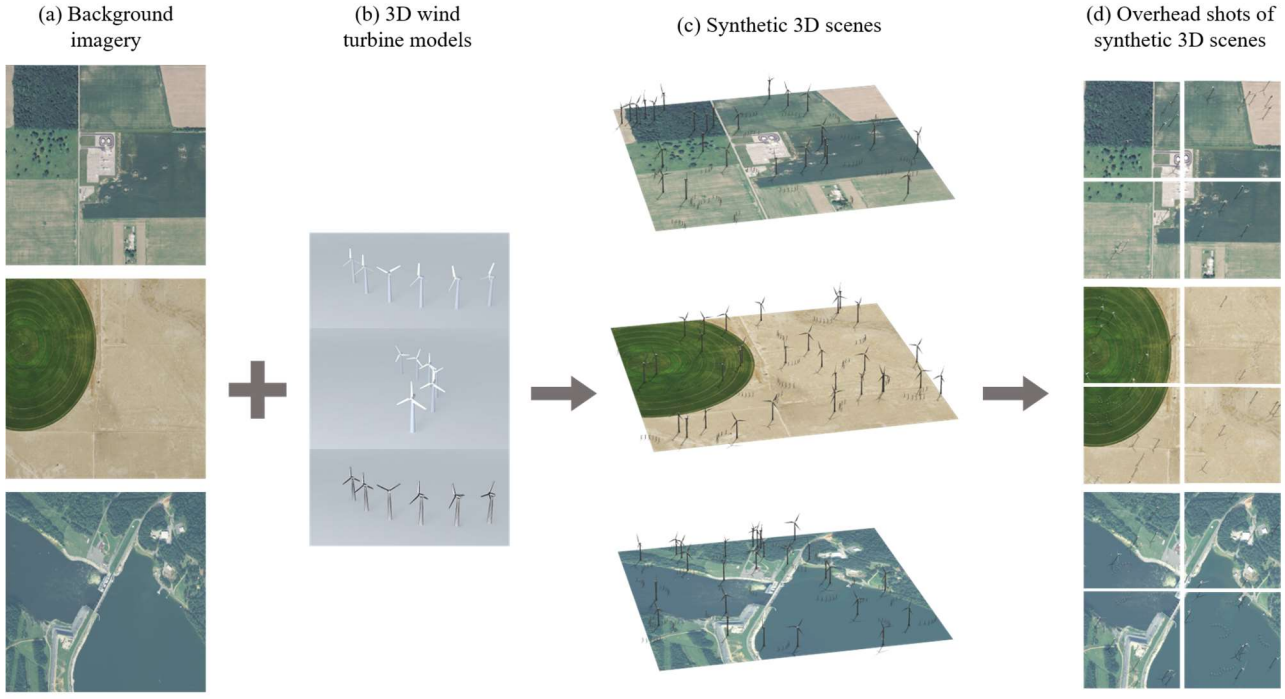

Figure 2. Synthetic imagery preparation: Real images from NAIP dataset (a) were used as the background to place 3D models of wind turbine objects (b) on to generate synthetic 3D scenes (c). Then overhead photos of generated 3D scenes were taken and were cropped into smaller training data patches (d) for further experiments.

contains wind turbines (Figure 2). To ensure the geographic diversity of the synthetic images in order to help the model generalize across diverse geographies, we randomly selected background images from these geographies: forest, farmland, grasslands, water, urban/suburban, mountains, and deserts (Figure 3). The selection of improbable geographies, such as water and forests, is designed to train the model to be less sensitive to the context. The model should learn detect a wind turbine because of the characteristics of object rather than the context the object is in. We also changed the lighting and the 3D models to make synthetic images more diverse on visual characteristics. In total, we created 2,796 synthetic images containing wind turbines. The modified dataset contained both the baseline training data and the synthetic images (4,133 images in total). The same testing set was used for all experiments. Both the baseline and the modified datasets were made public [8].

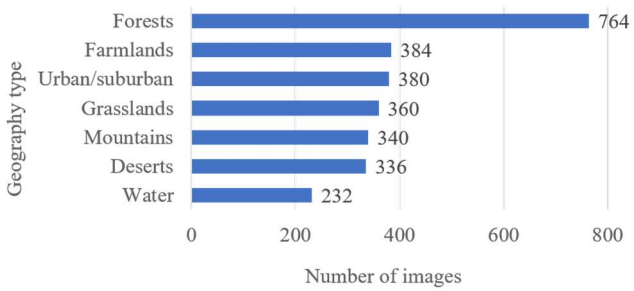

Figure 3. Number of images in the experimental dataset by geography type.

2.3. Wind turbine object detection model

In this work, the objects of interest are wind turbines as they are relatively rare and are an example of a type of energy infrastructure that would be valuable to have knowledge of for system-level planning and may be suggestive of the performance of similar classes of objects that are clearly visible in overhead imagery. In overhead imagery, the precise location of a wind turbine is the target of interest, so an object detection task is most appropriate for this application as opposed to scene classification. Object detection models take an image as input and output the coordinates of bounding boxes of the objects of interest within the image. In our experiments, we used YOLOv3, a well-cited CNN-based object detection architecture to train our models [3].

To evaluate our model's performance for each experimental condition, we calculated precision, recall, and average precision (AP, defined as the mean precision at a set of 11 equally spaced recall levels [10]) on the testing set. We also constructed precision-recall (PR) curves by varying the confidence threshold for making object detections.

It is worth noting that for this task, and many rare object identification tasks, recall tends to be the metric of greatest interest. Consider the task of finding every wind turbine in a country of 9 million square kilometers. Having to look everywhere would be incredibly time consuming. Recall quantifies what fraction of the objects that actually exist were found. We cannot remediate low recall manually without significant effort. Precision, on the other hand, is the fraction of detected objects that actually are what we classified them as. Low precision means we have many false positives. For rare object detection, precision is less of a concern, because human inspection and evaluation of the and removal of false positives is possible. Therefore, we seek to maximize recall and can often sacrifice some precision in the process.

3. EXPERIMENTS AND RESULTS

3.1. Impact of adding synthetic imagery into training data

To explore how synthetic imagery might affect a model's object detection performance, we designed experiments to compare using only real images for training and using real images combined with synthetic images for training. In particular, we are interested in the situation when the algorithm is highly data constrained, so we focus our experiments on a small subset of the larger dataset (10%) and compare to the much less data constrained case of the full dataset.

We first further trained the pre-trained YOLOv3 model on training set of the full baseline dataset (real images containing wind turbines). For these experiments we trained for 300 epochs with a batch size of 10 to produce our baseline model. We then trained on the modified dataset (which included synthetic data). In the modified case, we trained the

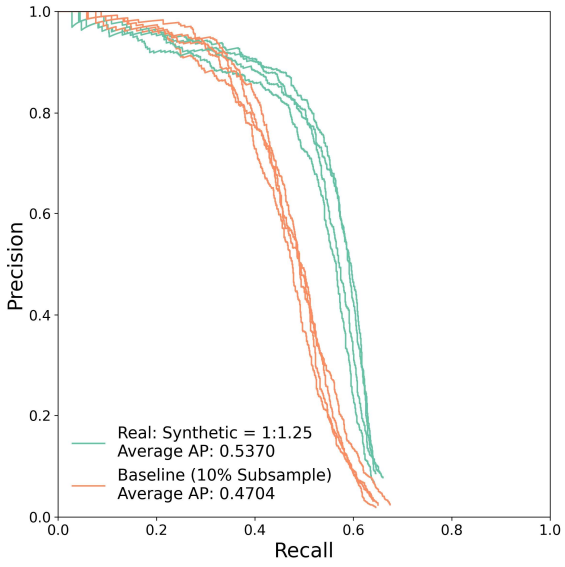

Figure 4. Precision-recall curves of models trained on with the 10% subsample of the baseline dataset and the modified dataset where the ratio of real images and synthetic images is 1:1.25.

model using the same model hyperparameter (pre-trained weights, number of epochs, batch size, etc.) except that the training dataset included synthetic images. After the two models were trained on the training set from each experimental condition, we tested them on the same testing set that consists of real images only and compared the performance. We repeated this process for the more data constrained case of a 10% random subsample of the training data. Note that for all experiments we repeated each model training 4 times to factor in variance in the model results and report averages for summary statistics such as AP.

Training on the full baseline dataset (without synthetic data) resulted in an AP of 0.7670 while the modified dataset (which contained the added synthetic data) yielded an AP of 0.7570, showing no improvement. We hypothesize this was because there were ample training samples included with adequate class variation to train the algorithm well, so the synthetic data did not further enhance performance, nor did it greatly reduce it.

We then randomly subsampled 10% (133 images) of all real training images to constrain the amount of real imagery available to mimic a highly-data constrained setting. This scenario will often be the case in many real-world application settings, where imagery is scarce. In this training-data-constrained case, the ratio of real:synthetic images was a particularly important variable to model performance so we varied it for a range of values when subsampling images from the synthetic data. All subsampling was done with the same procedure to maintain geographic balance. The results of this experiment were clear: the baseline model (without synthetic imagery) achieved an AP of 0.4704. When we added synthetic imagery to that baseline case, the best AP performance was 0.5370 (using the 1:1.25 real:synthetic ratio), a significant improvement. The full performance curves for this experiment are shown Figure 4.

4. CONCLUSIONS AND FUTURE WORK

This work evaluated an approach to create synthetic imagery to overcome the lack of annotated imagery for many object detection tasks by adding synthetic models of objects of interest to real imagery data. We designed experiments in which we trained YOLOv3 object detection models to detect wind turbines from overhead imagery datasets under two conditions: (1) with real images only and (2) with those same real images complemented with synthetic images created by adding artificial wind turbines to the real imagery. Our results suggest that adding synthetic images to augment the training data may yield performance improvements when the quantity of real imagery available for training is highly constrained—a common situation encountered in practice.

The synthetic imagery preparation process we piloted can be applied to imagery from any location, so for future work we will explore how this technique can help adapt an object detection method to new geographies by augmenting the training set with synthetic imagery from target geographies that have little to no labeled real imagery by utilizing more abundant unlabeled imagery augmented with synthetic objects such as wind turbines.

5. ACKNOWLEDGEMENTS

We acknowledge the financial support of the Duke University Data+ and Bass Connections programs as well as the Duke University Energy Initiative and the Rhodes Information Initiative. We also thank our collaborators Kathleen Chen, Eddy Lin, Wendy Zhang, Jose Moscoso, and Matt Robbins.

6. REFERENCES

[1] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[2] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Cham, 2018, pp. 270–279, doi: 10.1007/978-3-030-01424-7_27.

[3] F. Kong, B. Huang, K. Bradbury, and J. Malof, "The Synthinel-1 dataset: a collection of high resolution synthetic overhead imagery for building segmentation," 2020, pp. 1814–1823, Accessed: Dec. 17, 2020. [Online]. Available: https://openaccess.thecvf.com/content_WACV_2020/html/Kong_The_Synthinel-1_dataset_a_collection_of_high_resolution_synthetic_overhead_WACV_2020_paper.html.

[4] O. US EPA, "Emissions & Generation Resource Integrated Database (eGRID)," *US EPA*, Jul. 27, 2020. https://www.epa.gov/egrid (accessed Jan. 20, 2021).

[5] Kyle Bradbury *et al.*, "Power Plant Satellite Imagery Dataset." figshare, Aug. 16, 2017, doi: 10.6084/m9.figshare.5307364.v1.

[6] "NAIP Imagery," *temp_FSA_02_Landing_InteriorPages*. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/ (accessed Dec. 17, 2020).

[7] "Advanced 3D City Design Software | ArcGIS CityEngine." https://www.esri.com/en-us/arcgis/products/arcgis-cityengine/overview (accessed Dec. 17, 2020).

[8] Tyler Feldman *et al.*, "Baseline Dataset and Synthetic Images." figshare, Jan. 22, 2021, doi: 10.6084/m9.figshare.13626377.v4.

[9] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *ArXiv180402767 Cs*, Apr. 2018, Accessed: Dec. 17, 2020. [Online]. Available: http://arxiv.org/abs/1804.02767.

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.