RarePlanes: Synthetic Data Takes Flight

Jacob Shermeyer[1], **Thomas Hossler**[2], **Adam Van Etten**[1], **Daniel Hogan**[1], **Ryan Lewis**[1], **and Daeil Kim**[2]

[1]In-Q-Tel - CosmiQ Works, [jshermeyer, avanetten, dhogan, rlewis]@iqt.org
[2]AI.Reverie, [thomas.hossler, daeil]@aireverie.com

Abstract

RarePlanes is a unique open-source machine learning dataset that incorporates both real and synthetically generated satellite imagery. The RarePlanes dataset specifically focuses on the value of synthetic data to aid computer vision algorithms in their ability to automatically detect aircraft and their attributes in satellite imagery. Although other synthetic/real combination datasets exist, RarePlanes is the largest openly-available very-high resolution dataset built to test the value of synthetic data from an overhead perspective. Previous research has shown that synthetic data can reduce the amount of real training data needed and potentially improve performance for many tasks in the computer vision domain. The real portion of the dataset consists of 253 Maxar WorldView-3 satellite scenes spanning 112 locations and 2,142 km^2 with 14,700 hand-annotated aircraft. The accompanying synthetic dataset is generated via AI.Reverie's novel simulation platform and features 50,000 synthetic satellite images with $\sim$ 630,000 aircraft annotations. Both the real and synthetically generated aircraft feature 10 fine grain attributes including: aircraft length, wingspan, wing-shape, wing-position, wingspan class, propulsion, number of engines, number of vertical-stabilizers, presence of canards, and aircraft role. Finally, we conduct extensive experiments to evaluate the real and synthetic datasets and compare performances. By doing so, we show the value of synthetic data for the task of detecting and classifying aircraft from an overhead perspective.

1 Introduction

Over the last decade, computer vision research and the development of new algorithms has been driven largely by permissively licensed open datasets. Datasets such as ImageNet [7], MSCOCO [31], and PASCALVOC [11] (among others) remain critical drivers for advancement. Convolutional neural networks (CNNs), currently the leading class of algorithms for most vision tasks [38, 64], require a large amount of annotated observations. However, the development of such datasets is often manually intensive, time-consuming, and costly to create. An alternative approach to manually annotating training data is to create computer generated images and annotations (referred to as synthetic data). After creating realistic 3D environments, one can then generate thousands of images at virtually no cost. Such data has been shown to be effective for augmenting and replacing real data, thus reducing the burden of dataset curation. Synthetic datasets continue to be developed and have been notably helpful in various domains including: autonomous driving [41–43, 15], optical flow [32, 41, 28], facial recognition [27, 8, 26], amodal analysis [23, 9] and domain adaptation [6, 24, 22, 52] (see Section 2.1 for further detail).

Although synthetic datasets continue to become more prevalent, no expansive permissively licensed synthetic datasets exist in the context of overhead observation. Overhead imagery presents unique challenges for computer vision models such as: the detection of small visually-heterogeneous objects, varying look angles or lighting conditions, and unique geographies. As such, creating synthetic

Preprint. Under review.

datasets from an overhead perspective is a significant challenge and simulators must attempt to closely mimic the complexities of a spaceborne or aerial sensor as well as the Earth's ever-changing conditions. For example, to create a large and heterogeneous synthetic dataset, one must account for each sensors varying spatial resolution, changes in sensor look angle, the time of day of collection, shadowing, and changes in illumination due to the sun's location relative to the sensor. Furthermore, the simulator must be able to account for other factors such as the ground appearance due to seasonal change, weather conditions, and varying geographies or biomes.

While synthetic datasets certainly have the potential to be beneficial, they require a paired real dataset with shared features to baseline performance and quantitatively test value. However, few permissively licensed overhead datasets [10, 57, 45] exist that focus on detection or segmentation tasks and feature very-high resolution real imagery from an overhead perspective. Overhead datasets remain one of the best avenues for developing new computer vision methods that can adapt to limited sensor resolution, variable look angles, and locate tightly grouped, cluttered objects. Such methods can extend beyond the overhead space and be helpful in other domains such as face-id, autonomous driving, and surveillance.

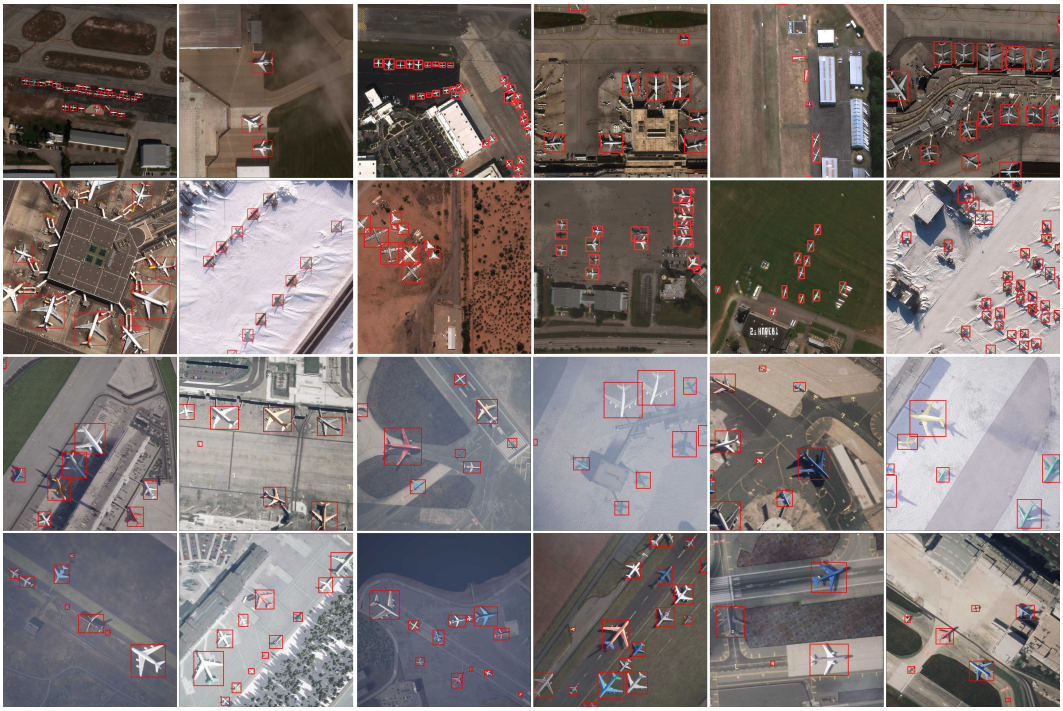


Figure 1: **Example of the real and synthetic datasets present in RarePlanes.** The top two rows feature the real Maxar WorldView-3 satellite imagery and the bottom two rows show the AI.Reverie synthetic data. The dataset features variable weather conditions, biomes, and ground surface types.

To address the limitations described above, we introduce the RarePlanes dataset. This dataset focuses on the detection of aircraft and their fine-grain attributes from an overhead perspective. It consists of both an expansive synthetic and real dataset. We use the AI.Reverie platform to develop realistic synthetic data based off of real world airports. The platform ingests real world metadata such as geospatial images to procedurally generate 3D environments of real world locations. The weather, time of collection, sunlight intensity, look angle, biome, and distribution of aircraft model are among the multiple parameters that the simulator can modify to create diverse and heterogeneous data. The synthetic portion of RarePlanes consists of 50,000 images and $\sim$ 630,000 annotations. The real portion consists of 253 Maxar WorldView-3 satellite images spanning 112 locations and 2,142km^2 with $\sim$ 14,700 hand annotated aircraft. Examples of the synthetic and real images are shown in Figure 1.

RarePlanes also provides fine-grain labels with 10 distinct aircraft attributes and 33 different sub-attribute choices labeled for each aircraft. These include: aircraft length, wingspan, wing-shape, wing-position, Federal Aviation Administration (FAA) wingspan class [17], propulsion, number of

engines, number of vertical-stabilizers, canards, and aircraft type or role. Although other overhead detection datasets exist [10, 57, 45, 29, 18, 55], no others have multiple fine-grain attributes that detail specific object features. Such fine-grain attributes have been particularly helpful for zero-shot learning applications [14] and enable end users to create diverse custom classes. Using these combined attributes, anywhere from 1 to 110 classes can be created for individual research purposes. The dataset is available for free download through Amazon Web Services' Open Data Program with download instructions and associated code available at `https://www.cosmiqworks.org/RarePlanes`.

Contributions

- An expansive real and synthetic overhead computer vision dataset focused on the detection of aircraft and their features.
- Annotations with fine-grain attributions that enable various CV tasks such as: detection, instance segmentation, or zero-shot learning.
- Extensive experiments to evaluate the real and synthetic datasets and compare performances. By doing so, we show the value of synthetic images for the task of detecting and classifying aircraft from an overhead perspective.

2 Related Work

RarePlanes sits at the intersection of three distinct computer vision dataset domains: synthetic datasets, geospatial datasets, and fine-grain attribution datasets. These three domains are cornerstones around which computer vision research has continued to rapidly advance and grow. We summarize the key characteristics of modern synthetic, geospatial, and attribute datasets in Table 1 and compare them to the RarePlanes dataset.

Table 1: **Comparison with other synthetic, attribute and overhead imagery datasets.** Our dataset has a similar scale as modern computer vision datasets and provides both a real and synthetic component. For SpaceNet (Buildings + Road Speed), xBD (Building Damage Scale), and RarePlanes we report the range of possible customizable classes that end-users can create using varieties of the dataset attributes.

Dataset	Gigapixels	Classes	Attributes	Labels	
				Real	Synthetic
SpaceNet [10, 57, 45]	100.1	1 to 8	1	859,982	0
xBD [18]	9.8	1 to 4	1	850,736	0
xView [29]	56.0	60	0	1,000,000	0
iSAID [55]	44.9	15	0	655,451	0
Cityscapes [5] + GTA [41]	537.5	30/19	0	210,179	510,4434
COCOA [65] + SAIL-VOS [23]	115.7	-/163	0	46,314	1,896,296
AWA2 [60]	24.7	50	85	37,322	0
CompCars [61]	86.1	1,716	13	136,726	0
RarePlanes (Ours)	187.1	1 to 110	10	14,707	629,551

2.1 Synthetic Datasets

Synthetic data has become prevalent across many computer vision domains and has shown value as a replacement for real data or to augment existing training datasets [42, 26, 43, 1, 37]. Many synthetic datasets focus on the autonomous driving domain; including the Synthia [43], GTA [41, 42], and vKITTI [15] datasets. These synthetic datasets are often paired with real-world data such as Cityscapes [5], CamVid [2], or KITTI [16] to benchmark the value of synthetic data. Other notable synthetic datasets such as SUNCG [47] or Matterport3D [3] focus on indoor scenes and include RGB-D data for depth estimation. Moreover, other datasets focus on addressing challenging occlusion (amodal) problems such as the expansive SAIL-VOS [23] and DYCE [9]. Finally, the Synthinel-1 [12] Dataset is the only other dataset that bridges the synthetic/geospatial domain. It features synthetic data from an overhead perspective with binary pixel masks of building footprints. Overall, combined synthetic and real datasets, similar to RarePlanes, have been helpful with several different tasks

including: enhancing object detection [49, 35, 37], semantic segmentation [43, 19, 44], or instance segmentation performance [56, 1]. Furthermore, such datasets continue to inspire new domain adaptation (DA) techniques [6, 24, 22, 52, 49]. Such DA techniques could be particularly valuable for overhead applications as there remains a dearth of openly available training data and models trained on one location often do not generalize well to new areas.

2.2 Geospatial Datasets

Geospatial and very-high resolution remote sensing datasets have continued to draw increased interest due to their relevancy to many computer vision challenges. Such datasets contain lower resolution images with tiny, closely grouped objects with varying aspect ratios, arbitrary orientations and high annotation density. The lessons learned from such datasets continue to inspire new computer vision approaches related to detection [62, 54, 51], segmentation [25], super-resolution [46, 36], and even bridges to natural language processing [50]. Some notable datasets include SpaceNet [10, 57, 45] and xBD [18], which focus on foundational mapping and instance/semantic segmentation for problems such as building footprint and road network extraction or building damage assessment. Others such as xView [29], A large-scale dataset for object detection in aerial images (DOTA) [59] and A Large-scale Dataset for Instance Segmentation in Aerial Images (iSAID) [55] focus on overhead object detection or instance segmentation, featuring multiple classes of different object types. The Functional Map of the World (FMOW) [4] dataset centers on the task of classification of smaller image chips from an overhead perspective. RarePlanes builds upon these existing datasets and contributes both synthetic and real data. Furthermore, RarePlanes adds 10 unique object attributes, which enable customizable classes, as well as three annotation styles per object (Bounding Box, Diamond Polygon, and Full-Instance (Synthetic Only)).

2.3 Fine-Grain Attribute Datasets

Many datasets focus on identifying general objects in imagery, however, several others take an alternative approach and label unique attributes of each object. As previously stated, RarePlanes features 10 attributes and 33 sub-attributes. Such attribution has been particularly valuable for constructing new zero-shot learning methods and algorithms [14]. The Comprehensive Cars [61] dataset is similar to RarePlanes and features attribute labels of 5 car attributes and 8 car-parts, as well as different look angles of vehicles. Several other similar datasets [60, 13, 53, 34, 63] feature multiple classes with extensive ranges in attributes; most of which are geared toward zero-shot learning research.

3 The RarePlanes Dataset and Statistics

Figure 2: **Three annotation styles within RarePlanes.** The RarePlanes synthetic dataset features three annotation styles including: 'Bounding Box' (left), 'Diamond Polygon' (center), and 'Full Instance Segmentation' (synthetic only) (right). Diamond annotations allow both wingspan and length to be calculated for each aircraft, as well as orientation.

3.1 Annotations, Features, and Attributes

The RarePlanes dataset contains 14,707 real and 629,551 synthetic annotations of aircraft. Each aircraft is labeled in a diamond style with the nose, left-wing, tail, and right-wing being labeled in successive order (Figure 2). This annotation style has the advantage of being: simplistic, easily reproducible, convertible to a bounding box, and ensures that aircraft are consistently annotated (other hand-annotated formats can often lead to imprecise labeling). Furthermore, this annotation style

enables the calculation of aircraft length and wingspan by measuring between the first annotation node to the third and from the second to the fourth. We employ a professional labeling service to produce high-quality annotations for the real portion of the dataset. Two rounds of quality control are included in the process, a first one by the professional service and a second by the authors.

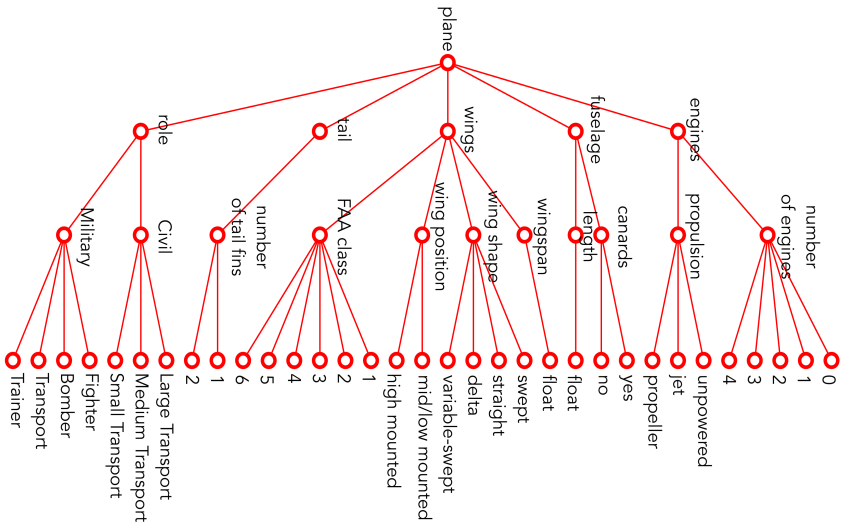


Figure 3: **The 5 features, 10 attributes, and 33 sub-attributes contained in the RarePlanes dataset.** The dataset and associated codebase (`https://github.com/aireveries/RarePlanes`) enables users to create custom classes using groupings of these attributes.

After each aircraft is annotated in the diamond format, an expert geospatial team labels aircraft features. The features include attributes of aircraft **wings**, **engines**, **fuselage**, **tail**, and **role** (Figure 3). We ultimately chose these attributes as they were visually distinctive from an overhead perspective and have been shown to be helpful in aiding to visually identifying the type or make of aircraft [40].

- ***Engines:*** we label the **Number of Engines:** ('0' to '4') and the **Type of Propulsion:** ('unpowered', 'jet', 'propeller').
- ***Fuselage:*** We label aircraft **Length in Meters:** ('float') and if the plane has **Canards:** ('yes' or 'no'). Canards are small fore-wings that are added to planes to increase maneuverability or reduce the load/airflow on the main wing.
- ***Wings:*** We label aircraft **Wing Shape:** ('straight', 'swept', 'delta', and 'variable-swept'), **Wing Position:** ('high mounted' and 'mid/low mounted'), **Wingspan in Meters:** ('float'), and the **FAA Aircraft Design Group Wingspan Class:** [17] ('1' to '6') which determines which airports can accommodate different sized aircraft. Examples of wing-shape and position can be seen in figure 4.
- ***Tail:*** We label the **Number of Vertical Stabilizers:** ('1' or '2') or tail fins that a plane possesses.
- ***Role:*** After labeling each attribute, we then use these attributes to classify the **Role or Type:** of an aircraft into seven unique classes. These include: 'Civil Transport/Utility' ('Small', 'Medium, and 'Large' based upon wingspan), 'Military Transport/Utility/AWAC', 'Military Bomber', 'Military Fighter/Interceptor/Attack', and 'Military Trainer'. Further detail on role definitions and can be found in the RarePlanes User Guide, hosted here: `https://www.cosmiqworks.org/RarePlanes`

3.2 Real Imagery and Locations

All electro-optical imagery is provided by the Maxar Worldview-3 satellite with a maximum ground sample distance (GSD) of 0.31 to 0.39 meters depending upon sensor look-angle. The dataset consists of 253 unique scenes, spanning $2{,}142\,\mathrm{km}^2$ with 112 locations in 22 countries. Locations were chosen by performing a stratified random sampling of OpenStreetMap aerodromes of area $\geq 1\,\mathrm{km}^2$ across

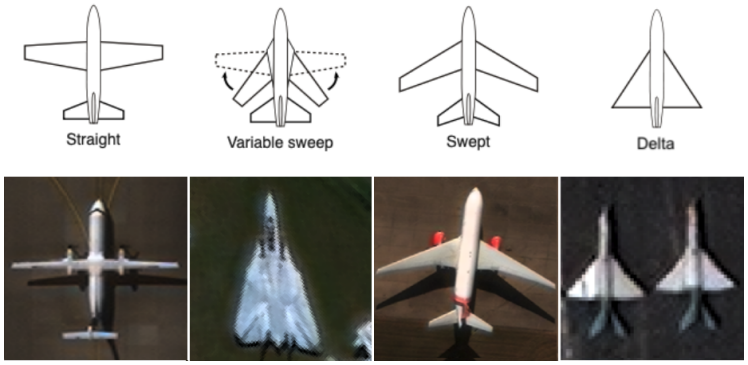


Figure 4: **Wing Shapes [48] present in the RarePlanes dataset.** Note that the two left-most aircraft feature 'high-mounted' wings, with the two right-most aircraft featuring 'mid/low mounted' wings.

the US and Europe using the Köppen climate zone as the stratification layer. We stratify by climate to increase seasonal diversity and geographic heterogeneity. Seven additional locations were manually chosen as they overlap with preexisting datasets [57, 10], and we considered further revisits over these locations to potentially have additional value. We then chose individual satellite scenes by attempting to select scenes from different seasons for each location. Many locations have several scenes taken at different points in time, which may enable future investigation on the value of annotating the same areas using multiple images. The imagery is collected from variable look-angles (3.2 to $29.6°$), target azimuth angles (1.8 to $359.7°$), and sun elevation angles (10.7 to $79.0°$). Imagery is collected from all four seasons, with scenes featuring instances of cloud cover (12.6%), snow (9.1%) and clear skies (78.3%). Combined together, this leads to high variability in illumination, shadowing, and lighting conditions. Consequently, the dataset should help to improve generalizability to new areas. Finally, background surfaces are quite diverse with grass, dirt, concrete, and asphalt surface types.

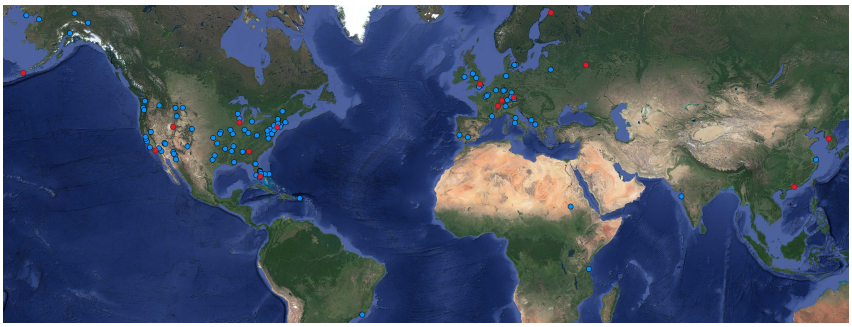


Figure 5: **RarePlanes dataset locations.** The dataset features 112 real (blue points) and 15 synthetic locations (red points). Atlanta, Miami, and Salt Lake City feature both real and synthetic data.

The collection is composed of three different sets of data with different spatial resolutions: one panchromatic band ($0.31 - 0.39$m), eight multi-spectral (coastal to NIR ($400 - 954\mu$m)) bands ($1.24 - 1.56$m), and three RGB ($448 - 692\mu$m) pan-sharpened bands ($0.31 - 0.39$m). Each data product is atmospherically compensated to surface-reflectance values by Maxar's AComp [33] and ortho-rectified using the SRTM DEM. RGB data is also converted to 8-bit. Areas containing non-valid imagery are set to 0. We distribute both 512×512 pixel tiles (20% overlap) that contain aircraft as well as as the full images, cropped to the extent of the area of annotation.

3.3 Synthetic Imagery

All synthetic data is created via the AI.Reverie simulator software. The synthetic dataset contains $629{,}551$ annotations of aircraft across $50{,}000$ images and 15 distinct locations, simulating a total area of $9331.2\,\mathrm{km}^2$. Each image features a simulated GSD of 0.3 meters and is collected from variable look-angles ranging between 5.0 to $30.0°$ off-nadir. The imagery is evenly split across 5 distinct biomes including: 'Alpine', 'Arctic', 'Temperate Evergreen Forests', 'Grasslands', and 'Tundra'. The

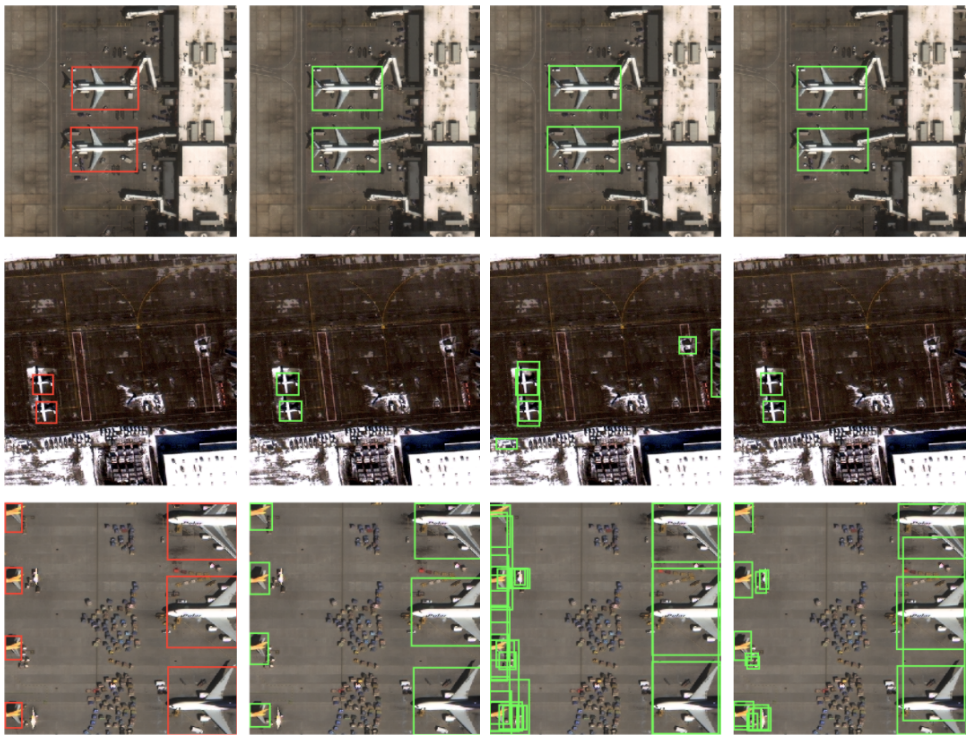
biome parameter controls the type of vegetation, its density, as well as the ground textures. Four unique weather conditions are also evenly distributed across the dataset including: 'Overcast', 'Clear Sky', 'Snow', and 'Rain'. Other parameters include the sunlight intensity, weather intenstiy, and the time of the day. Ultimately, this produces an expansive heterogeneous dataset with a wide variety of backgrounds. We believe that this dataset will be helpful in improving model generalizability to new areas and developing new algorithmic approaches that could move beyond aircraft detection.

4 Experiments, Results, and Discussion

In this section, we validate the synthetic dataset by running three experiments for two tasks: object detection and instance segmentation. For each task, we train a benchmark network on three subsets of data: on the real data only, on the synthetic data only, and perform a fine tuning experiment training on the synthetic data and then a portion ($\sim 10\%$) of the real dataset. Each experiment is validated on the test real dataset and the results are shown Table 2. We ran these experiments for two attributes: aircraft (detection of an aircraft without classifying it) and civil role.

4.1 Training and Testing Splits

For the real world data, given the size of the raw satellite scenes, we adopted a tiling approach. Each scene has been cut into 512x512 tiles containing at least one aircraft. Furthermore, we ensure that the training and test split contains at least one satellite scene per country. As the dataset contains multiple satellite images captured over the same location, an airport can appear in both splits at different points in time. Moreover, we created a subset of the real training split for the fine tuning experiments. This subset contains roughly 10 percent of the images of the training split, created by drawing a 10% random sample of image tiles by location. For the synthetic data, images were randomly split into a training set containing $45{,}000$ images and a testing set of $5{,}000$ images, which we used primarily for cross-validation. Note those results are not reported here.



(a) (b) (c) (d)

Figure 6: **Example of aircraft detection results**. (a) ground truth, (b) model trained real dataset (c) model trained on synthetic dataset (d) model fine tuned on real subset.

4.2 Implementation

In our experiments, we used a Resnet-50 [21] and FPN [30] as the backbone for the Faster R-CNN [39] detection network. A similar backbone was used for the Mask R-CNN [20] instance segmentation network. Backbones are pre-trained using ImageNet [7]weights and all experiments are conducted with the Detectron2 framework [58], using the default configurations for each network. The network was optimized with Stochastic Gradient Descent (SGD) using a learning rate of 0.001, weight decay of 0.0001 and a momentum of 0.9. Additionally, we used a linear warmup period over 1K iterations. We maintain a consistent learning rate for the fine tuning experiments. We found that decreasing the learning rate or freezing some of the layers in the backbone did not improve performance. The networks were trained on a NVIDIA Tesla V100 GPU with 12GB memory. Each network was trained until convergence, which was reached after around 60K iterations. We also applied basic pixel level augmentations, such as blurring and modifying the contrast or the brightness. Finally, we performed random cropping (512x512) when training on the synthetic dataset.

4.3 Results and Discussion

We evaluated our network performances using the COCO average precision (AP) metric. Table 2 reports the average precision for each class as well as the mAP, mAP50, and average recall (AR). Qualitative results are shown in Figure 6.

Table 2: **Results of the object detection and segmentation experiments.** We report models performance trained on the real dataset (Real) and the synthetic dataset (Synth.) as well as the fine tuning experiment (FT) using only 10% of the real training dataset. We show the results of the single class experiments ('aircraft') and the three classes experiment: small (C_S), medium (C_M), and large (C_L) civil transport aircraft. Performance is evaluated using the mean average precision (mAP) (IOU@[0.5:0.95]), the mAP50 (IOU@0.5) and the average recall (AR) metrics, as well as the class APs when applicable. For the Mask R-CNN instance segmentation experiments, we only report the segmentation AP. Each value reported is an average of 5 runs. The standard deviations for mAP, mAP50, and AR are also indicated.

network	attribute	dataset	C_S	C_M	C_L	mAP	mAP50	AR
Faster R-CNN	aircraft	Real	N/A	N/A	N/A	73.32 (0.34)	96.80 (0.02)	77.16 (0.21)
	aircraft	Synth.	N/A	N/A	N/A	54.86 (0.25)	87.03 (0.53)	60.67 (0.27)
	aircraft	FT	N/A	N/A	N/A	69.16 (0.69)	95.29 (0.41)	73.03 (0.57)
	role	Real	66.68	70.26	67.68	68.21 (0.4)	92.16 (0.23)	75.39 (0.40)
	role	Synth.	27.70	37.09	42.85	35.88 (2.26)	59.09 (2.9)	53.82 (1.28)
	role	FT	56.73	66.05	66.52	63.10 (0.78)	89.15 (0.22)	71.06 (0.75)
Mask R-CNN	aircraft	Real	N/A	N/A	N/A	73.67 (0.17)	96.81 (0.03)	76.46 (0.20)
	aircraft	Synth.	N/A	N/A	N/A	56.28 (0.46)	87.54 (0.69)	60.71 (0.51)
	aircraft	FT	N/A	N/A	N/A	70.51 (0.34)	94.73 (0.03)	73.72 (0.26)
	role	Real	65.60	72.13	70.97	69.57 (0.47)	91.89 (0.55)	76.16 (0.30)
	role	Synth.	29.12	41.78	47.47	39.46 (3.20)	62.31 (4.51)	57.33 (1.96)
	role	FT	58.96	70.02	72.33	67.11 (0.46)	90.03 (0.52)	74.40 (0.58)

In the first set of experiments, we focused on the performance of the synthetic dataset only. As expected, we observe a drop in performances when training on the synthetic data only, due to the domain gap between the real and synthetic datasets. We observe that the model trained on the synthetic dataset tends to mislabel clutter or nearby objects as aircraft, as shown in Figure 6. Additionally, snow patches, ground markings, airport vehicles are sometimes detected as aircraft. This leads to a significantly lower AP (55% to 75% of the real AP) when models are trained on the synthetic dataset only. However, the AR is not as sensitive to the domain gap (70% to 80% of the real AR), meaning that the majority of aircraft are still detected when only the synthetic dataset is used. Similarly, we observe that the drop in AP50 is also lower relative to the AP metric. Ultimately, the AP50 metric may be more informative as we are most interested in accurately counting aircraft, rather than how well they are localized.

Most importantly, when a small subset ($\sim 10\%$) of real data is added for fine tuning, we observe a significant gain in mAP, leading to similar performance to the models trained on the real dataset

only. We hypothesize that the synthetic data helps to build a prior model for aircraft detection and eases transfer learning, thus greatly reducing the need for annotated real data. In Figure 6, we see how fine tuning on the real subset removes some of the false positive predictions versus training on the synthetic dataset only. However, the false positive detection rate still remains slightly higher compared to training on the entire real training set. It's important to note that the goal of these experiments is to define a baseline for future experimentation for other algorithms to improve upon, particularly within the area of domain adaptation.

Acknowledgment

The authors thank the whole AI Reverie team for making the creation of the synthetic dataset possible. We would especially like to thank Danny Gillies and Natasha Ruiz for their devoted help.

References

[1] H. A. Alhaija, S. K. Mustikovela, L. Mescheder, A. Geiger, and C. Rother. Augmented reality meets deep learning for car instance segmentation in urban scenes. In *British machine vision conference*, volume 1, page 2, 2017.

[2] G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2):88–97, 2009.

[3] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *International Conference on 3D Vision (3DV)*, 2017.

[4] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee. Functional Map of the World. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018.

[5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[6] G. Csurka. Domain adaptation for visual applications: A comprehensive survey, 2017.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[8] B. Egger, W. A. P. Smith, A. Tewari, S. Wuhrer, M. Zollhoefer, T. Beeler, F. Bernard, T. Bolkart, A. Kortylewski, S. Romdhani, C. Theobalt, V. Blanz, and T. Vetter. 3d morphable face models – past, present and future, 2019.

[9] K. Ehsani, R. Mottaghi, and A. Farhadi. Segan: Segmenting and generating the invisible. In *CVPR*, 2018.

[10] A. V. Etten, D. Lindenbaum, and T. M. Bacastow. Spacenet: A remote sensing dataset and challenge series, 2018.

[11] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[12] K. B. J. M. Fanjie Kong, Bohao Huang. The synthinel-1 dataset: a collection of high resolution synthetic overhead imagery for building segmentation. In *2020 Winter Conference on Applications of Computer Vision (WACV)*, 2020.

[13] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1778–1785. IEEE, 2009.

[14] Y. Fu, T. Xiang, Y.-G. Jiang, X. Xue, L. Sigal, and S. Gong. Recent advances in zero-shot recognition, 2017.

[15] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig. Virtual worlds as proxy for multi-object tracking analysis. In *CVPR*, 2016.

[16] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[17] S. Gudmundsson. *General aviation aircraft design: Applied Methods and Procedures*. Butterworth-Heinemann, 2013.

[18] R. Gupta, B. Goodman, N. Patel, R. Hosfelt, S. Sajeev, E. Heim, J. Doshi, K. Lucas, H. Choset, and M. Gaston. Creating xbd: A dataset for assessing building damage from satellite imagery. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[19] A. Handa, V. Patraucean, V. Badrinarayanan, S. Stent, and R. Cipolla. Understanding real world indoor scenes with synthetic data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4077–4085, 2016.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn, 2017.

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[22] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. A. Efros, and T. Darrell. Cycada: Cycle-consistent adversarial domain adaptation, 2017.

[23] Y.-T. Hu, H.-S. Chen, K. Hui, J.-B. Huang, and A. G. Schwing. SAIL-VOS: Semantic Amodal Instance Level Video Object Segmentation – A Synthetic Dataset and Baselines. In *Proc. CVPR*, 2019.

[24] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz. Multimodal unsupervised image-to-image translation. *Lecture Notes in Computer Science*, page 179–196, 2018.

[25] V. Iglovikov and A. Shvets. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation, 2018.

[26] A. Kortylewski, B. Egger, A. Morel-Forster, A. Schneider, T. Gerig, C. Blumer, C. Reyneke, and T. Vetter. Can synthetic faces undo the damage of dataset bias to face recognition and facial landmark detection?, 2018.

[27] A. Kortylewski, A. Schneider, T. Gerig, B. Egger, A. Morel-Forster, and T. Vetter. Training deep face recognition systems with synthetic data. *arXiv preprint arXiv:1802.05891*, 2018.

[28] P. Krähenbühl. Free supervision from video games. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2955–2964, 2018.

[29] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric, Y. Bulatov, and B. McCord. xView: Objects in context in overhead imagery. *CoRR*, abs/1802.07856, 2018.

[30] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection, 2016.

[31] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[32] N. Mayer, E. Ilg, P. Fischer, C. Hazirbas, D. Cremers, A. Dosovitskiy, and T. Brox. What makes good synthetic training data for learning disparity and optical flow estimation? *International Journal of Computer Vision*, 126(9):942–960, Apr 2018.

[33] F. Pacifici, N. Longbotham, and W. J. Emery. The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6241–6256, Oct 2014.

[34] G. Patterson, C. Xu, H. Su, and J. Hays. The sun attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision*, 108(1-2):59–81, 2014.

[35] X. Peng, B. Sun, K. Ali, and K. Saenko. Learning deep object detectors from 3d models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1278–1286, 2015.

[36] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao. Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network. *Remote Sensing*, 12(9):1432, 2020.

[37] P. S. Rajpura, H. Bojinov, and R. S. Hegde. Object detection using deep cnns trained on synthetic images, 2017.

[38] W. Rawat and Z. Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.

[39] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2015.

[40] D. Rendall. *Jane's Aircraft Recognition Guide*. HarperCollins, 1996.

[41] S. R. Richter, Z. Hayder, and V. Koltun. Playing for benchmarks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2232–2241, 2017.

[42] S. R. Richter, V. Vineet, S. Roth, and V. Koltun. Playing for data: Ground truth from computer games. In *European conference on computer vision*, pages 102–118. Springer, 2016.

[43] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016.

[44] S. Sankaranarayanan, Y. Balaji, A. Jain, S. Nam Lim, and R. Chellappa. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3752–3761, 2018.

[45] J. Shermeyer, D. Hogan, J. Brown, A. V. Etten, N. Weir, F. Pacifici, R. Haensch, A. Bastidas, S. Soenen, T. Bacastow, and R. Lewis. Spacenet 6: Multi-sensor all weather mapping dataset, 2020.

[46] J. Shermeyer and A. Van Etten. The effects of super-resolution on object detection performance in satellite imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[47] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser. Semantic scene completion from a single depth image. *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[48] Steelpillow. Aircraft — Wikipedia, the free encyclopedia. `https://commons.wikimedia.org/wiki/User:Steelpillow/Aircraft`, 2020. [Online; accessed 30-April-2020].

[49] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[50] B. Uzkent, E. Sheehan, C. Meng, Z. Tang, M. Burke, D. Lobell, and S. Ermon. Learning to interpret satellite images in global scale using wikipedia, 2019.

[51] A. Van Etten. You only look twice: Rapid multi-scale object detection in satellite imagery. *arXiv preprint arXiv:1805.09512*, 2018.

[52] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Perez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019.

[53] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

[54] C. Wang, X. Bai, S. Wang, J. Zhou, and P. Ren. Multiscale visual attention networks for object detection in vhr remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 16(2):310–314, 2019.

[55] S. Waqas Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. Shahbaz Khan, F. Zhu, L. Shao, G.-S. Xia, and X. Bai. isaid: A large-scale dataset for instance segmentation in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 28–37, 2019.

[56] D. Ward, P. Moghadam, and N. Hudson. Deep leaf segmentation using synthetic data. *arXiv preprint arXiv:1807.10931*, 2018.

[57] N. Weir, D. Lindenbaum, A. Bastidas, A. V. Etten, S. McPherson, J. Shermeyer, V. Kumar, and H. Tang. Spacenet mvoi: a multi-view overhead imagery dataset. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 992–1001, 2019.

[58] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.

[59] G.-S. Xia, X. Bai, Z. Z. Jian Ding, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang. DOTA: A Large-scale Dataset for Object Detection in Aerial Images. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Nov. 2017.

[60] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2251–2265, Sep 2019.

[61] L. Yang, P. Luo, C. Change Loy, and X. Tang. A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3973–3981, 2015.

[62] X. Yang, J. Yang, J. Yan, Y. Zhang, T. Zhang, Z. Guo, S. Xian, and K. Fu. Scrdet: Towards more robust detection for small, cluttered and rotated objects, 2018.

[63] B. Zhao, Y. Fu, R. Liang, J. Wu, Y. Wang, and Y. Wang. A large-scale attribute dataset for zero-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[64] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.

[65] Y. Zhu, Y. Tian, D. Metaxas, and P. Dollár. Semantic amodal segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1464–1472, 2017.