IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323289

VEHSAT: A LARGE-SCALE DATASET FOR VEHICLE DETECTION IN SATELLITE IMAGES

Sébastien Drouyer[⋆]

[⋆]CMLA, ENS Cachan, CNRS, Université Paris-Saclay, France

ABSTRACT

There has been this last decade a major improvement in earth observation imagery, both in terms of resolution and revisit rate. This technological leap comes with an increasing demand for detecting objects in satellite images. In this context, vehicle detection generates a great deal of interest in the scientific and industrial community. Indeed, detecting vehicles can have a range of applications from traffic monitoring to parking lots occupancy rate estimation. This interest is demonstrated in the wide range of public vehicle datasets that already exist. However, these datasets rely on the use of aerial images, acquired from planes, that are different in many aspects from satellite images. Machine learning algorithms trained on these datasets are therefore likely to underperform on satellite images. Therefore, we introduce a large-scale dataset for VEHicle detection on SATellite images (VehSat). To this end, we collected 4544 crops of satellite images, from 4 different satellites and on 8 different areas. In total, 36851 vehicles were annotated in various contexts and resolutions. To build a baseline for vehicle detection in satellite images, we also evaluate state-of-the-art object detection algorithms on VehSat. Results show that, although performing better than human annotators on crowdsourcing platforms, machine learning algorithms still have a lot of room for improvement due to the challenging nature of the dataset.

Index Terms— vehicle detection, satellite image, dataset, deep-learning

1. INTRODUCTION

Automated vehicle detection on satellite images have a range of applications in the private and public sectors. It allows a cheap, recurrent and full-scale traffic monitoring that is indicative of the economic activity [1,2]. It can be used to keep track of the volume business of retailers by measuring the occupancy rate of their parking lots [3–5]. It could also assist governments in environmental missions such as countering illegal logging by monitoring trucks near forest areas.

Satellite images are particularly interesting when a frequent monitoring of an area is needed. Although aerial images are of much better quality, they require that a plane flies over the area of interest for each acquisition, making the cost prohibitive for frequent monitoring. On the other hand, satellite images can be bought for a fraction of the cost from an increasing amount of suppliers, and recent actors such as Planet allow to obtain images on a daily basis.

In order to allow automated vehicle detection, several datasets already exist containing images and ground truths [6–10]. However, these datasets mainly use aerial images and not satellite images. The nature of these images are very different, so a detection algorithm will likely underperform if trained on aerial images and then applied on satellite images. The key differences are the Ground Sample Distance (around 20cm for aerial images, between 31cm and 80cm on the satellite images of our dataset), the number of bands (aerial images are RGB images, satellite images can contain more than 8 bands), the value range (0-255 for aerial images, 0-65535 for satellite images) and the acquisition system (classical cameras for aerial images and push broom scanners for satellite images) that can create unique artefact on satellite images such as the pushbroom effect (see Figure 1).

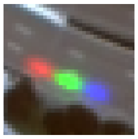


Fig. 1: Image of a moving vehicle acquired using SkySat.

Moreover, the aspect of satellite images varies depending on the satellite being used and the atmospheric conditions of the area of interest during the acquisition. This can affect the signal to noise ratio as well as the general color balance.

To advance vehicle detection research on satellite images, this paper introduces VehSat. VehSat is a large scale dataset that contains crops of satellite images associated with manually annotated ground truths. We collected 4544 crops of images from different satellites - SkySat, Pleiades, World-View 2 and World-View 3 - and annotated, using crowdsourcing and a manual post-checking process, 36851 vehicles. In order to avoid overfitting to a specific region, our images have been extracted from several areas: Atlanta (USA), Boston

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. Work partly financed by Office of Naval research grant N00014-17-1-2552, DGA Astrid project ”filmer la Terre” n°ANR-17-ASTR-0013-01 and Kayrros, Inc. We thank Carlo de Franchis (Kayrros) and Jean-Michel Morel (ENS Paris-Saclay) for valuable suggestions. We also thank ESA for providing Pleiades images.

(USA), Khartoum (Sudan), Las Vegas (USA), Lima (Peru), Paris (France), Rome (Italy), Shanghai (China).

To our knowledge, VehSat is the first public annotated vehicle dataset on satellite images that combines several sources and several regions. It can be used to develop and evaluate vehicle detectors in satellite images. We plan to frequently update VehSat by adding new annotations and new regions.

2. ANNOTATION OF VEHSAT

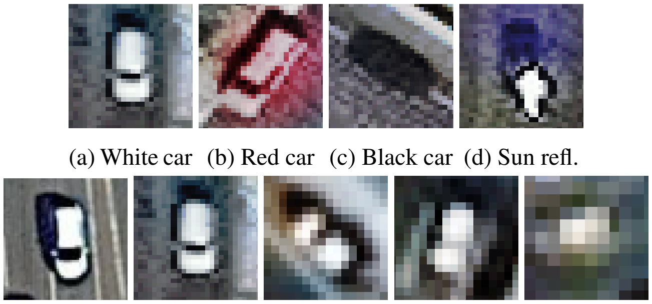
2.1. Vehicle definition and aspect

We annotate an object as a vehicle if it is a car or a non -rail motorized vehicle larger than a car (pickups, trucks, buses). We decided not to annotate motorcycles as they are difficult to spot even on images with the best resolution.

A clear challenge of the dataset is the large variability of the aspect of vehicles. Even if we limit our observation to cars, whose dimensions are the most stable, the way they appear in images can differ substantially.

White and colored cars are the easiest to spot as their windshields are generally discernible (figures 2a, 2b). Black cars are more challenging to detect as they generally appear as a black homogeneous area, making them hardly distinguishable from shadows. See Figure 2c.

Other effects, such as sun reflections, can appear (see Figure 2d). The acquisition's Ground Sample Distance also has a clear impact on the way cars appear, even after resizing the images so that cars have a similar dimension in pixels. See Figure 2.

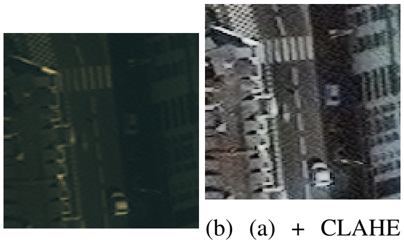


(a) White car (b) Red car (c) Black car (d) Sun refl.

(e) DOTA (f) WV-3 (g) WV-2 (h) Pleiades (i) SkySat

Fig. 2: First row: examples of vehicles on World-View 3 images (PAN-sharpened). Sun refl.: Sun reflection on black car. Second row: Examples of white cars from aerial and satellite images. DOTA: aerial image of vehicle taken from the DOTA [9] dataset. WV-3: WorldView-3. WV-2: WorldView-2.

There are also regional variations. The length of cars depends on the area of interest, probably due to cultural and economic factors. For instance, there are more limousines in Las Vegas than in Lima. In cities where buildings are high, the shadow can cover roads, making vehicles less visible. See Figure 3.



(a) Original crop (b) (a) + CLAHE [11]

Fig. 3: Image of Paris acquired using WorldView-3. Some vehicles are covered by shadows, making them difficult to spot. Crop in (a) is displayed so that each channel is comprised between 0 and 255. CLAHE [11] has been applied in (b) so that the contrast is enhanced on shadowed areas.

2.2. Images collection

WorldView-2 and WorldView-3 images have been collected from the Spacenet dataset [12], SkySat images from the Planet website [13] and Pleiades images from the TPM Online Catalogue of the European Space Agency. Images are then divided into different non-overlapping crops for facilitating the annotation process. The sizes of the crops are chosen so that the area covered is at least 70×70 meters. As the GSD of each satellite is different, the size of crops measured in pixels is also different. See table 1.

2.3. Annotation method

On this dataset, each car has been annotated by a bounding box centered on the vehicle. Annotations are obtained in two steps. Each crop is annotated independently. We first use crowdsourcing to obtain a first version. The annotations are then manually corrected and standardized by the authors in the second step.

2.4. Dataset splits

For machine learning purposes, we randomly split the dataset into three nonoverlapping sets: 60% in the training set, 20% in the validation set and 20% in the test set. We will publicly provide all the crops associated with the ground truth for the training set and the validation set. We will only provide the images for the testing set. An online benchmark tool will allow users to upload their prediction for the testing set so that they can be evaluated independently.

AOI	**Satellite model**	**GSD**	**Nb. bands**	**Crops size**	**Nb. crops**	**Area (km^2)**	**Nb. instances**
Atlanta (USA)	WorldView-2	46	8	154	796	3.99	6578
Boston (USA)	SkySat	80	4	100	607	3.88	8545
Khartoum (Sudan)	WorldView-3	31	8	256	600	3.78	2873
Las Vegas (USA)	WorldView-3	31	8	256	420	2.65	1643
Lima (Peru)	Pleiades	50	4	154	453	2.69	4950
Paris (France)	WorldView-3	31	8	256	600	3.78	4315
Rome (Italy)	Pleiades	50	4	154	468	2.77	5478
Shanghai (China)	WorldView-3	31	8	256	600	3.78	2469
TOTAL	N/A	N/A	N/A	N/A	4544	27.32	36851

Table 1: Coverage per Area Of Interest. GSD is indicated in cm/px. Crops size is indicated in px.

3. DATASET DETAILS

The dataset contains in total of 4544 crops, covering 27.32 square kilometers and containing 36851 instances of vehicles. See table 1. WorldView-3 crops account for a little bit less than 50% of the dataset (2220 / 4544). It is worth noting that the density of vehicles is very variable across the dataset: there are 2202 vehicles per square kilometers in Boston (USA), whereas there are 760 vehicles per square kilometers in Khartoum (Sudan). There are between 0 to 124 cars per crop, with an average of 7.5 and a standard deviation of 10.1.

4. EVALUATION

4.1. Methology

As the size of crops and the number of bands they contain can vary depending on the source satellite, the range of pixels is much higher (uint16), and the size of instances can also be very small (approximately 6×2 pixels on SkySat images), crops need to be standardized so that common machine learning algorithms can exploit them. The standardization is done in three steps. First, only the Red, Green and Blue channels are retained from these crops. CLAHE [11] is then applied on the crops, allowing to get a uint8 image where the contrast has been enhanced on shadowed areas. Finally, crops are resized to 512×512 pixels using bilinear interpolation, allowing to standardize the size of instances in dimensions that can be usable on detection algorithms.

4.2. Benchmark

We trained and evaluated YOLOv2 [14] and the Faster R-CNN [15] algorithms as they are the state of the art methods both on general object detection and on object detection on aerial images [9]. We also split the dataset per satellite and trained the Faster R-CNN algorithm on each subset to check how a specialized detection algorithm would perform. Results are shown in Table 2.

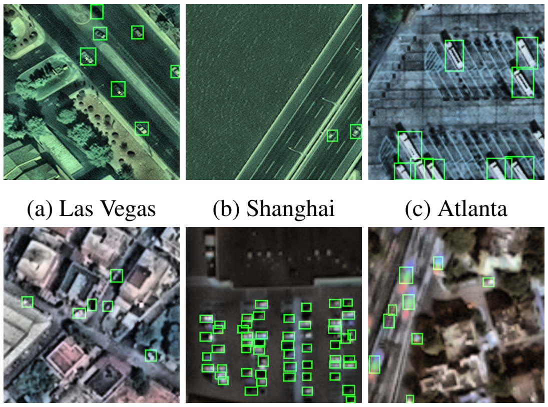
Average Precisions per satellite are shown in table 3.

Method	**Precision**	**Recall**	**AP**
FRCNN [15]	43.9%	73.4%	63.6%
FRCNN-Spe	29.1%	69.6%	56.5%
CS	91.2%	47.8%	43.9%
YOLOv2 [14]	31.4%	32.8%	28.8%

Table 2: Benchmark of state of the art detection algorithms on VehSat. FRCNN: Faster R-CNN algorithm. FRCNN-Spe: Faster R-CNN algorithm specialized for each satellite. CS: Crowdsourcing. AP: average precision.

Method	**WV3**	**WV2**	**Pleiades**	**SkySat**
FRCNN [15]	75.8%	70.2%	51.9%	48.3%
FRCNN-Spe	66.6%	66.8%	46.2%	43.0%
CS	70.3%	52.8%	31.2%	23.7%
YOLOv2 [14]	39.6%	47.1%	13.4%	9.1%

Table 3: Comparison of the Average Precision per source satellite for each method. FRCNN: Faster R-CNN algorithm. FRCNN-Spe: Faster R-CNN algorithm specialized for each satellite. CS: Crowdsourcing.



(a) Las Vegas (b) Shanghai (c) Atlanta

(d) Rome (e) Boston (f) Boston

Fig. 4: Examples of detections obtained using the Faster R-CNN algorithm (non specialized).

4.3. Discussion

Examples of detections obtained using the Faster R-CNN algorithm is shown in Figure 4.

Overall, the performance of state of the art algorithms is far from perfect on the VehSat dataset, probably due to its very challenging nature. According to the Average Precision and the recall criteria, the Faster R-CNN algorithm seems to be the best performing, even better than the initial annotations obtained by crowdsourcing.

Training a specialized Faster R-CNN for each satellite doesn't seem to give better results than training the algorithm on all images, Average Precision being even slightly higher on the second approach. This is an interesting outcome because it means that Faster R-CNN has been trained in a way so that it is robust to changes in resolution, at least for the 4 satellites being tested.

As it is the case for the annotations obtained using crowdsourcing, performance of all detection algorithms decreases as the GSD of the image increases.

5. CONCLUSION

We created VehSat, the first large-scale dataset for vehicle detection on satellite images. It contains images from four different satellites (WorldView-3, WorldView-2, Pleiades, SkySat) with different resolutions and different number of bands. Images were acquired on 8 different areas: Atlanta (USA), Boston (USA), Khartoum (Sudan), Lima (Peru), Paris (France), Rome (Italy), Shanghai (China), Las Vegas (USA).

The dataset, the benchmark system, and source code of detection algorithms will be available from:

`http://www.vehsat.com`

VehSat is a very challenging dataset on several aspects. First, it contains instances that can be as small as a dozen of pixels, that can be partly covered by trees, buildings, or shadows. The density of instances in each crop can vary widely between 0 and 124. The resolution of images varies from 31 centimeters per pixel to 80 centimeters per pixel. Some images contain 4 bands, others contain 8 bands. The range of value of each pixel is comprised between 0 and 65535 with very high values on sun reflections, instead of 0 and 255 in standard images.

These challenging aspects are confirmed in the benchmark we produced using state of the art algorithms, where performance metrics - though better in some cases than annotations given by crowdsourcing - are far from achieving perfect scores.

In addition to providing valuable data for vehicle detection algorithms, we believe that VehSat poses interesting questions on general object detection, such as the detection of small objects, sub-pixel detection, orientation detection, detection in multi-resolutions datasets and detection on images with different number of channels.

6. REFERENCES

[1] Floris J van Ruth, "Traffic intensity as indicator of regional economic activity," Tech. Rep., Citeseer.

[2] Sébastien Drouyer and Carlo de Franchis, "Highway traffic monitoring on medium resolution satellite images," in *International Geoscience and Remote Sensing Symposium*, 2019.

[3] Crawford J, "Beyond supply and demand: making the invisible hand visible," 2016.

[4] Sébastien Drouyer and Carlo de Franchis, "Parking occupancy estimation on sentinel-1 images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020.

[5] Sébastien Drouyer, "Parking occupancy estimation on planetscope satellite images," in *International Geoscience and Remote Sensing Symposium*, 2020.

[6] Franklin Tanner, Brian Colder, Craig Pullen, David Heagy, Michael Eppolito, Veronica Carlan, Carsten Oertel, and Phil Sallee, "Overhead imagery research data set—an annotated data library & tools to aid in the development of computer vision algorithms," in *2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009)*. IEEE, 2009, pp. 1–8.

[7] Sébastien Razakarivony and Frédéric Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016.

[8] T Nathan Mundhenk, Goran Konjevod, Wesam A Sakla, and Kofi Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," in *European Conference on Computer Vision*. Springer, 2016, pp. 785–800.

[9] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *CVPR*, 2018.

[10] Gong Cheng, Junwei Han, Peicheng Zhou, and Lei Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 98, pp. 119–132, 2014.

[11] Karel Zuiderveld, "Graphics gems iv," chapter Contrast Limited Adaptive Histogram Equalization, pp. 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[12] "Spacenet on amazon web services (aws). "datasets." the spacenet catalog.," `https://spacenetchallenge.github.io/datasets/datasetHomePage.html`, Accessed: 2018-11-08.

[13] "Skysat samples," `https://info.planet.com/download-free-high-resolution-skysat-image-samples/`, Accessed: 2018-11-08.

[14] Joseph Redmon and Ali Farhadi, "Yolo9000: Better, faster, stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, 2017.

[15] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.