

Fine-Grain Uncommon Object Detection from Satellite Images

L. Lee

MIT Lincoln Laboratory

244 Wood St., Lexington, MA, USA
leel@ll.mit.edu

B. Smith

MIT Lincoln Laboratory

244 Wood St., Lexington, MA, USA
benjamin.smith@ll.mit.edu

T. Chen

MIT Lincoln Laboratory

244 Wood St., Lexington, MA, USA
tina.chen@ll.mit.edu

Abstract—The ever increasing amount of earth observing satellite images is a vast treasure trove of interesting objects. We address the topic of object detection from satellite images in cases where the object is rarely observed and hence there is very little availability of images to support training classifiers. Unlike objects observed on the ground, there is no equivalent ImageNet with labeled data for objects as seen from satellite or aerial platform sensors that could be used to train classifiers. In addition, we focus on specific uncommon objects with very limited observations. To overcome the lack of training data, we built a near-class object detector and verified the uncommon object detection using images from different domains. We demonstrate the performance of our uncommon object detector and show a high detection rate in satellite images.^{1,2}

I. INTRODUCTION

The wide availability of high-resolution imaging satellite sensors results in massive amount of detailed images of earth. These commercial satellite images span in resolutions from 0.7 meter pixels (Quickbird) to 0.25 meter pixels (Worldview 3). Many classes of large objects are readily identifiable with human eyes from these high-resolution images, such as vehicles, boats, airplanes, buildings and trees. Automating the analysis of aerial/satellite images to extract high-level content understanding will significantly increase the utilization of overhead images because the vast quantity of images collected from satellite sensors makes manual analysis prohibitively expensive. Hyperspectral images have been applied in automated classification tasks for trees, agricultural land use [11], and water body detection [2]. Detection of common object classes such as buildings [12], [10], baseball fields [7], and boats [15] from overhead images have also been studied. We focus on the problem of detecting specific fine-grained objects where the targets are highly uncommon in the test image domain of satellite images and hence there is a severe lack of training images available for training fine-grained object detectors. We approach the challenge by building a near-class detector followed by a verification step using cross-domain image data of the specific uncommon target object.

Automatic object detection/classification has been an active area of computer vision research, with significant advances made after the establishment of common data sets such as

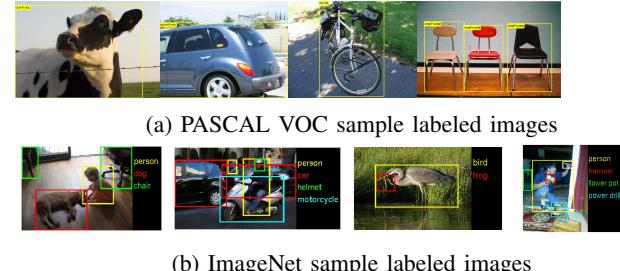


Fig. 1: Examples of labeled images from PASCAL VOC and ImageNet image databases

the PASCAL Visual Object Classification database and the ImageNet databases [5], [13]. The power of these common databases cannot be underestimated. They have allowed researchers to leverage large databases of labeled images to experiment with and improve algorithms for training object classifiers [6], [9]. In the case of ImageNet, it contained 1000 object categories and millions of images. Figure 1 shows examples of images and labels from PASCAL VOC and ImageNet. All images from common image databases used to train and test object classifiers are collected from ground based sensors, such as handheld cameras, with a ground-based viewing angle to objects. Aerial and satellite images of earth-bound objects are very different in appearance from ground-based views of the same objects because of the dramatic viewing angle differences. Everingham *et al.*

citeEveringham14 show that even much smaller viewing angle anomalies dramatically reduce object detection/classification rate. In addition to the challenge of viewing angle differences, there is no existing database of labeled objects in the domain of overhead/satellite images, which forces one to build a database of object classes. In the case of uncommon objects, such images may not be available at all. Our contribution in this paper is to provide a framework to tackle the lack of training data in the uncommon object detection problem. Our approach involves training a detector of a more common near-class or super-class objects, then verifying the presence of the uncommon object using images from a number of image domains, exclusive of satellite and aerial image, by transforming those verification images into one with uniform characteristics.

The remainder of this paper is laid out as follows: Section II references prior work on object detection; Section III details

¹This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract number FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

²Copyright:978-1-4673-9558-8/15/\$31.00 2015 IEEE

our approach to the problem of uncommon object detection, by building a hierarchical classifier that detects the super-class and verifies the specific class with various cross-domain data sources; Section IV describes our experiments with our uncommon object detection pipeline and presents our results.

II. RELATED WORK

The field of object detection from images has made tremendous advances in the years since the start of contests such as PASCAL VOC and ImageNet LSVRC. We will discuss key works in the following three areas: general object detection from optical images, object detection from satellite/aerial images, and training data augmentation for object detection.

Object detection from optical images generally involve the following steps: transform images to feature spaces, then train classifiers using labeled data and apply to test data. Prior to 2012, one of the most successful feature transforms for detection applications was the Histogram of Gradient (HoG) descriptor by Dalal and Triggs [3]. HoG descriptors are computed by dividing images into overlapping blocks and taking histograms of the gradient image for each block to form a feature set. The HoG descriptor was originally applied to pedestrian detection from images and showed much improved performance over previously existing image descriptors such as the Haar wavelet transform. Almost all image feature descriptors had been hand designed until 2012, when Krizhevsky *et al.* [9] applied deep convolution neural nets to visual object detection. The authors applied deep learning techniques to unlabeled images from ImageNet database to arrive at a hierarchical feature representation for object detection. They won the ILSVRC challenge with a significant margin against the second place finisher.

While object detection from satellite images has had less attention than object detection from ground based images, there has been some work on the topic. Three such examples are baseball diamond detection by Ke *et al.* [7], building detection by Lin *et al.* [10] and Mueller *et al.* [12], and ship detection by Yokoya *et al.* [15]. Ke *et al.* used Haar wavelet features coupled with an Adaboost classifier to train a cascade of features for detection of baseball diamonds. Mueller *et al.* used region segmentation and shape priors to detect buildings, while Lin *et al.* used edges and shape priors not only to detect buildings but also to infer their 3D shape. Yokoya *et al.* used sparse representations to find features for image patches depicting parts of objects and applied them to the problem of ship detection in coastal environments.

Training classifiers for object detection often requires sufficient number of images of the target objects. In cases where there are not enough real (and labeled) images of the target object, synthetic images have been used to augment training data sets. Rozantsev *et al.* [1] showed that object detector performance improves with the addition of 3D rendered synthetic images to real data, as long as the synthetic data was generated to match the characteristics of real data. Sun and Saenko [14] showed that without matching image statistics of 3D rendered images to real images, it is possible to apply adaptation learning to classifiers and still achieve good detection results.

Our focus on this work is in the detection of specific uncommon objects from aerial/satellite images. We chose to

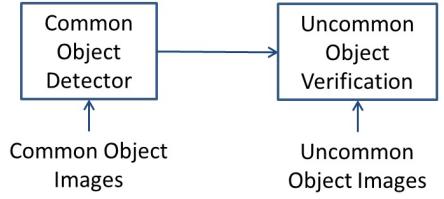


Fig. 2: Uncommon object detection pipeline

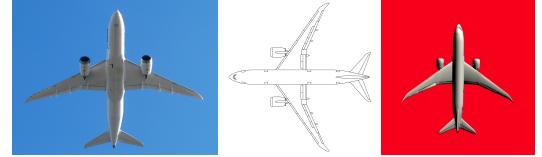


Fig. 3: Models of overhead view of the Dreamliner plane from left to right: a planform diagram, an underbelly view taken from the ground, and a CAD model rendering.

use simple feature representations such as HoG rather than apply deep learning to extract better features. We also augment our object model with cross-domain data sources and synthetic data.

III. UNCOMMON OBJECT DETECTION

Most object detection pipelines have the following elements: a labeled training image data set, an image feature transform, and a classifier trained using the transformed features of training images. Our uncommon object detection pipeline (Fig. 2) starts by finding a commonly observable near-class object on which we can build a detector, followed by a verification step to confirm the uncommon object using cross-domain image data (non-satellite or aerial image).

We illustrate our approach with a specific object, the Dreamliner airplane. Our goal is to detect and localize all instances of the Dreamliner airplane in any satellite images. One common practice when faced with collecting images of objects not commonly present in existing databases is to perform a textural search for images using relevant terms relying on common sense labeling of crowd-provided data. Conventional textual search with terms "Dreamliner satellite images" on either Google or Bing image search show no relevant satellite images of the Dreamliner airplane. They do return a few interesting images, such as a planform engineering drawing of the Dreamliner, a ground-collected image of the underbelly of the plane in almost planform view, and a CAD model [3]. Dreamliner is a subclass of commercial airliner, which is a common object category easily observable from overhead views of most major airports. We utilize the similarity in appearance of the Dreamliner jet to other commercial jets to build a near-class detector for jetliners, then verify the presence of the specific uncommon object, the Dreamliner, using alternate data sources.

A. Near-class Object Detector

The approach taken for the detection of airplanes using commercially available satellite imagery in this work is detailed in this section. Commercial aircraft have a distinct, rigid

shape that can be taken advantage of using a number of state-of-the-art computer vision techniques. In order to detect airplanes of different shapes, orientations and sizes the detection method must be robust to scale and orientation changes and have the capacity to generalize the details of aircraft shape. To accomplish this task an object detection algorithm was developed using a linear Support Vector Machine to learn Histogram of Oriented Gradients (HoG) feature descriptors on overhead imagery, including satellite and aerial data.

1) Features and Learning: Histograms of Oriented Gradients (HoG) [3] is an image feature descriptor that is based on the local binning of edges in an image based on orientation and spatial location. This feature descriptor provides robustness to small shape variations that are typically seen in different types of large commercial aircraft. Additionally, HoG descriptors provide some invariance to the small changes in view point differences typically seen in electro-optical satellite imagery. Instead of using the 36-dimensional HoG descriptor of Dalal and Triggs [3], we used the reduced dimensional set of features detailed in [6]. This yields a 31-dimensional HoG descriptor that includes both contrast-sensitive and -insensitive features. These HoG descriptors are calculated for each location in the image using a spatial block size of 8×8 pixels. A linear support vector machine (SVM) trained using *dlib* [8] is used to learn the optimal HoG filter for detection of commercial airplanes in electro-optical satellite data. To provide invariance to large rotations of the object (>10 degrees) a HoG filter is learned for every 15 degrees of rotation of the airplanes, resulting in 24 oriented detectors. In addition to rotational invariance, invariance to scale is provided by operating the learned filter across an image pyramid that is down-sampled by $5/6$ at each level. The detection window that provided the best detection performance was 80×80 pixels with no margin added around the airplanes. This provides support to detect an airplane in any possible rotation and at any size within the set of image resolutions used in this work.

2) Training: To train the system described above, a number of airplane images were taken from freely available sources such as Bing Maps and Google Earth. In total, 110 airplane images were used as positive examples. Each of these positive training examples was manually rotated such that all aircraft appeared in the same rotation for training, which allowed each orientation to be created by rotating the positive training examples by the appropriate angle. All positive training examples were then resized to the appropriate detection window size chosen for this work of 80×80 pixels. Negative examples were collected at random from overhead images that did not contain airplanes.

B. Object Verification

Using the location, scale, and orientation estimate from the output of the near-class object detector, the airplane detections can be verified against Dreamliner information in several different domains to determine which detections are of the correct object (Fig. 5). Specifically, we use the Dreamliner planform engineering drawing, gradient from an image of the underbelly of the plane, and gradient from a rendering of a CAD model of the aircraft. The planform image and the CAD rendering were both from nadir view as would be seen from a satellite or aerial platform. The underbelly view of the Dreamliner taken from

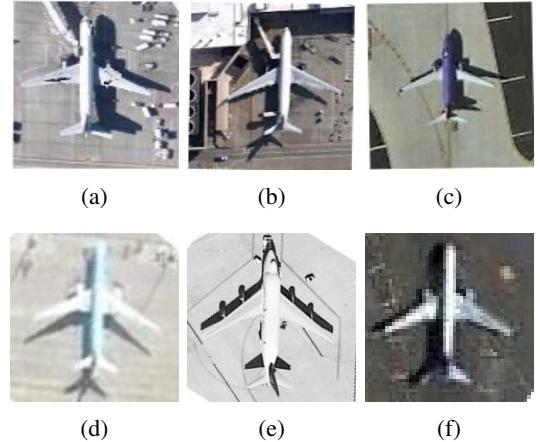


Fig. 4: Example of images used in training a commercial jet detector

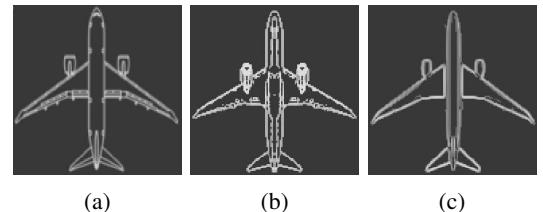


Fig. 5: Dreamliner information from different domains is used for plane verification: (a) planform drawing, (b) underbelly view image gradient and (c) CAD model gradient

the ground had perspective distortion, which was corrected using publicly available information regarding aspect ratios of its wing span versus fuselage length.

1) Verification Algorithm: The images of the near-class object detections are first rotationally aligned according to the angle estimated by the oriented detector. A better angular estimate is determined by a convolution between the image gradients of the chips and the gradient of each of the three Dreamliner representations at angles between -8 and 8 degrees, in increments of 1 degree. The angle that yields the maximum response determines the correct plane alignment. Finally, Dreamliner aircrafts are identified by the correlation score at the correct orientation (Fig.6). We have experimentally determined that the best performing input to the correlation process is a truncated gradient image where gradient values below a certain threshold are set to zero. The two questions that remained to be answered are 1, the threshold to truncate the gradient inputs to the correlation process, and 2, the classifier

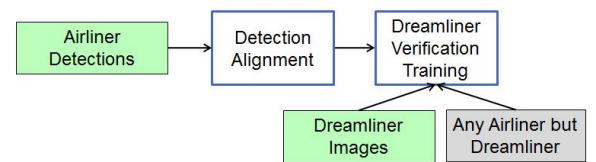


Fig. 6: Process to classify airliner detections as Dreamliner.

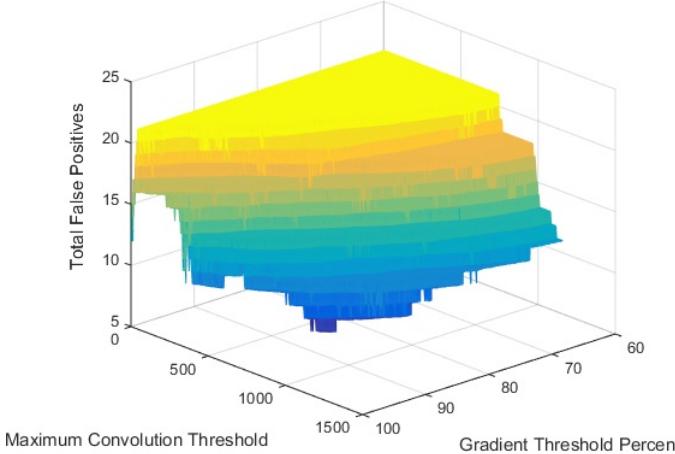


Fig. 7: Total number of false positives in the threshold training set at each threshold combination. Note that the maximum number of false positives is 21, which is equal to the number of non-Dreamliner airplane detected in the threshold training set images.

boundary of the correlation score for the Dreamliner verifier.

2) *Training:* To avoid arbitrarily picking thresholds for truncating gradient values and for the correlation score of positive Dreamliner verification, we learn these values using overhead images with and without Dreamliner presence. The positive training set with Dreamliner presence came from a Google Maps image of Charleston, South Carolina, site of the Dreamliner assembly plant. The negative data was randomly chosen from any major airport with a significant presence of commercial jetliners but no Dreamliners. We exhaustively sampled the joint threshold space and evaluated the verification performance for false positive Dreamliner identification (Fig. 7). Since we are primarily interested in detecting a rare object, we only consider threshold combinations that detect all of the Dreamliner instances in the training set images. Among the threshold values that yield the lowest number of false positives, we consider several threshold combinations, such as the two extremes (highest convolution threshold with the lowest gradient percentile and the lowest convolution threshold with the highest gradient percentile) and a convolution and gradient percentile combination in the center of the valid threshold options. The results do not show significant differences, so we use the centrally located point. These thresholds are then applied to test images without further modification. Note that while we used one actual Dreamliner overhead image to determine the thresholds, this can be achieved using another commonly observable similar targets.

IV. EXPERIMENTS AND RESULTS

We test our uncommon object detection approach of building a near-class detector using commonly observable object followed by a verification step specific to the uncommon object using publicly available aerial images and tasked satellite images. In addition, we implemented a comparison approach where we trained a Dreamliner detector in the same process

Total Detections	Recall ³	Precision
46	1.000	0.978

TABLE I: Results of near-class object detection on test image

as our commercial jet detector, using CAD rendering of the Dreamliner model with varying lighting changes with and without random background inserted, as training data. The CAD model trained Dreamliner detector failed to detect planes in our test data and was eliminated for any further comparison.

A. Data

Experiments were conducted on Bing Maps and Google Maps images at 0.5 m resolution of airports in Boston, MA, Charleston, SC, and Phoenix, AZ, as well as three orthorectified satellite images of the Boston area collected by the Pleiades satellite at 0.5 m resolution covering 100 km². Charleston, SC, is the location of Dreamliner assembly facility and hence had many examples of Dreamliners visible on Google Maps. The Charleston airport provided the positive data, while Phoenix airport—which had no Dreamliner presence—provided the negative data to train the threshold for the verification process. Bing Maps image of Boston Logan airport, covering 1.5 km², and three Pleiades satellite images of Boston and the surrounding area, each covering 100 km², were used as testing images.

B. Results

Table I summarizes the results of the near-class object detection on the testing imagery shown in Figure 8. As can be seen in Figure 8, the system does an excellent job of correctly classifying large commercial jet aircraft. There are a noticeable number of airplanes that are not detected by the system. While these appear to be missed detections, they are not. The system was trained to detect commercial jet aircraft for the purpose of detecting the specific 787 Dreamliner, and as such, the system does not detect small personal aircraft or aircraft that deviate significantly in shape from commercial jets. The system achieves a 100% true detection rate with a minimal number of false alarms, 0.01 per square kilometer.

Table II summarizes the results of the Dreamliner verification on the data sets described above. The Dreamliner planform image yielded the fewest false positives in the test sets. Though the CAD model and the underbelly image had a higher number of false positives, the results indicate that in the absence of an available planform image, these two alternatives can still provide object verification after using a near-class detector. Each of the four test images contained one Dreamliner, samples of which are shown in (Fig. 9). Some examples of the false positives are in Figure 10.

C. Discussion

Our Dreamliner detector trained using 3D rendering of a CAD model performed poorly in the test images. The potential causes could be that systematic biases in the background and in image statistics such as lack of noise in rendered images.

³Does not include small private aircraft or 'straight wing', non-commercial jet aircraft.

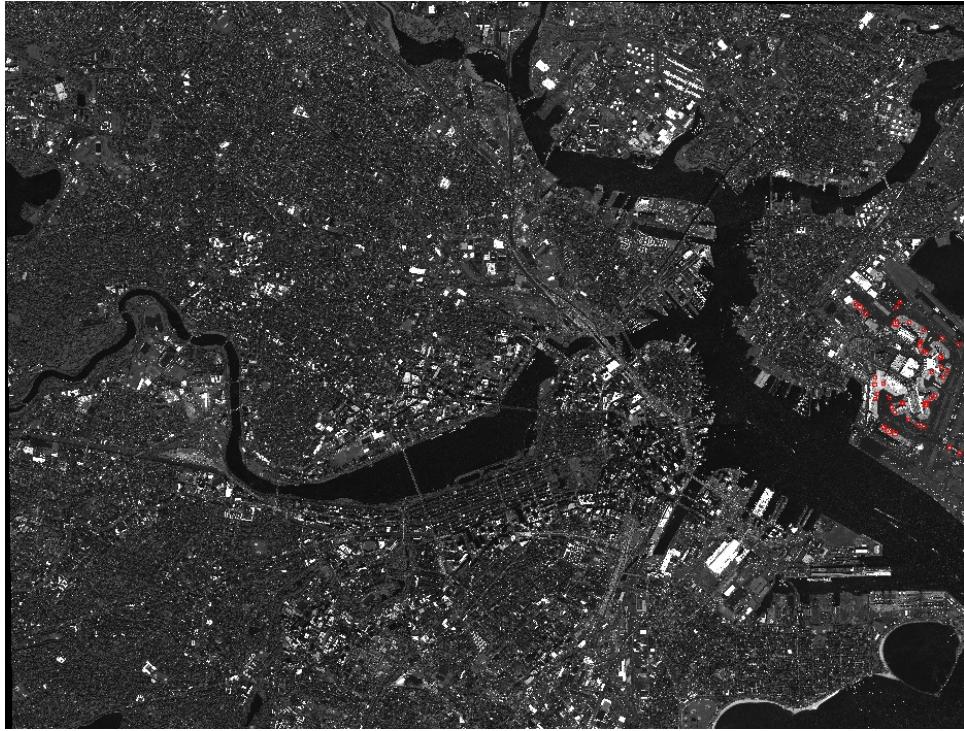


Fig. 8: Airplane detection example on testing image from Pleiades satellite. Red boxes mark the detections.

Test Data	Planform Results		Underbelly Image Results		CAD Model Results	
	Dreamliners	False Positives	Dreamliners	False Positives	Dreamliners	False Positives
Logan (GoogleMaps)	1	4, 2.7/km ²	1	1, 0.7/km ²	1	3, 2/km ²
Boston 1	1	1, 0.01/km ²	1	6, 0.06/km ²	1	3, 0.03/km ²
Boston 2	1	0, 0/km ²	1	2, 0.02/km ²	1	6, , 0.06/km ²
Boston 3	1	2, 0.02/km ²	1	5, 0.05/km ²	1	6, , 0.06/km ²

TABLE II: Results of Dreamliner verification on near-class object detections using three data sources: planform diagram, belly view of the Dreamliner, and CAD model rendering. Per km² false positive identification rates are significantly worse for the Logan airport image because the data is a small sample covering only the airport, and hence contains a large number of potentially confusing targets.



Fig. 9: Samples of correctly verified Dreamliner images.

Improved training image synthesis as performed by [1] could increase detector performance. One could also apply adaptation learning as per Sun and Saenko [14] to improve detector performance. The false positive Dreamliner detections after the verification step could be reduced by weighting the gradient in the verification models by significance of the edge/gradient feature as compared to a common commercial jet model.

V. CONCLUSION

We presented an approach to detecting uncommon objects from satellite/aerial images. Uncommon objects present a

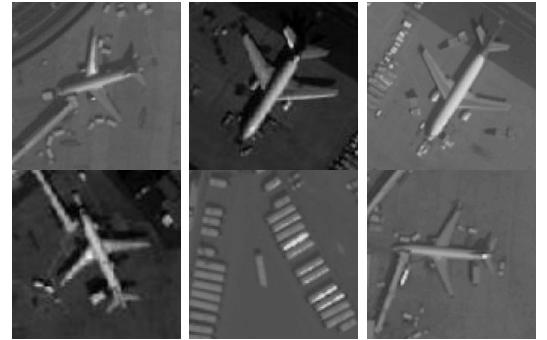


Fig. 10: Example of false positives from the Dreamliner verification step.

particular challenge in that little data is available to apply the traditional training process, which requires a large number of training images. Our approach leverages appearance similarity of the uncommon object to other commonly observable objects to build a near-class detector, followed by a verification step

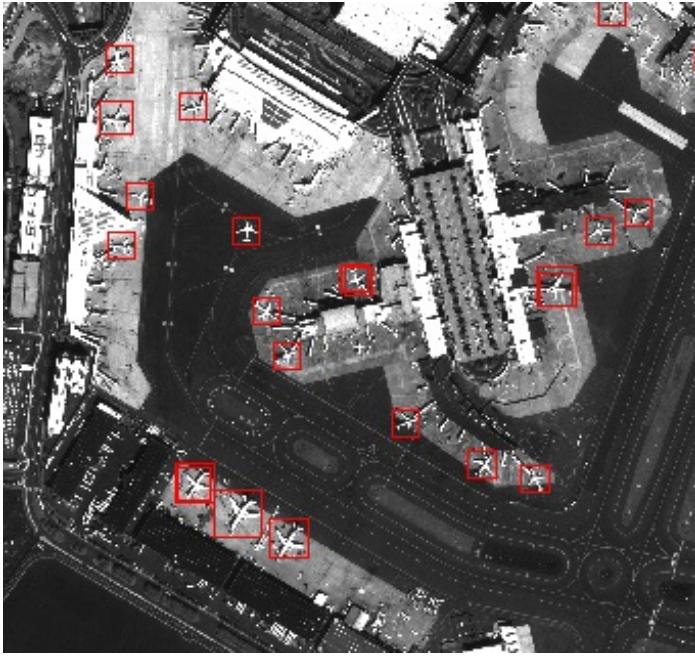


Fig. 11: Moderate zoom on airport region of testing image. Red boxed regions indicate airplane detections.



Fig. 12: High zoom view of airport region of testing image. Red boxed regions indicate airplane detections.

where we use non-aerial or satellite images, such as engineering diagrams, transformed ground-based images, and CAD model renderings transformed to a common framework. We demonstrated detection performance on satellite images and achieved high detection rate.

REFERENCES

- [1] V. L. A. Rozantsev and P. Fua. On rendering synthetic images for training an object detector. *Computer Vision and Image Understanding*, 2015.
- [2] M. Bochow, B. Heim, C. Rogass, I. Bartsch, K. Segl, and H. Kaufmann. Automatic detection and delineation of surface water bodies in airborne hyperspectral data. In *Proceedings of International Geospatial and Remote Sensing Symposium*, 2012.
- [3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detections. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [4] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge - a retrospective. *International Journal of Computer Vision*, 2014.
- [5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.
- [6] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Trans. on PAMI*, 32(9):1627–1645, 2010.
- [7] Y. Ke, J. Zhao, Z. Yuan, C. Qu, S. Han, Z. Zhang, X. Jiang, and G. Liang. A rapid object detection method for satellite image with large size. In *International Conference on Multimedia Information Networking and Security*, volume 1, pages 637–641, Nov 2009.
- [8] D. King. *Dlib*, 2009. <http://www.dlib.net/>.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, pages 1097–1105. Curran Associates, Inc., 2012.
- [10] C. Lin and R. Nevatia. Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, 1998.
- [11] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.*, 42(8), 2004.
- [12] S. Mueller and D. W. Zaum. Robust building detection in aerial images. In *Proceedings International Society for Photogrammetry and Remote Sensing, Workshop CMRT*, 2005.
- [13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge, 2014.
- [14] B. Sun and K. Saenko. From virtual to reality: Fast adaptation of virtual object detectors to real domains. In *British Machine Vision Conference*, 2014.
- [15] N. Yokoya and A. Iwasaki. Object localization based on sparse representation for remote sensing imagery. In *Proceedings of International Geospatial and Remote Sensing Symposium*, 2014.