

Object Detection with Satellite Imagery: using YOLO to Detect and Localize Objects in Aerial Satellite Imagery

Canyon Robins, Jaydon Krok, {crobins, jkrok}@stanford.edu Stanford University

Abstract

- At this very moment, there are just under 5000 satellites orbiting the planet. In 2018 we saw ar increase in the number of satellites from 2017 of
- This increase in satellites will inevitably bring with it an increased proliferation of satellite imagery. These images give us an incredible tool to solve certain problems such as resource allocation during disasters and the ability to more effectively monitor
- In our project we used the YOLO model to apply a deep learning solution to the problem of object detection using Aerial Satellite Imagery.

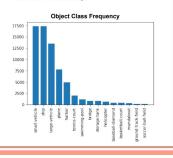
Results:

We were able to achieve mAP for our best class of 0.87, but our mAP over all classes was only 0.1336, which is less than desired performance. We were able to identify clear next steps, and our model architecture shows promise for stronger performance with future work.

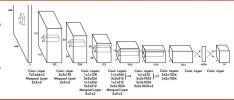
Data Source, Cleaning and Transformation

We used the DOTA satellite aerial images object detection dataset Contains 1869 aerial satellite images ranging in size from 800 x 800 to 4000 x 4000

- There are 15 object classes labeled and boxed in each image
- · Some objects are labelled as difficult
- Some appeared rarely, and some objects were very small
- We removed the difficult examples and wrote code to crop and flip images to improve accuracies on those classes where this was an issue e.g. detecting boats
- The resulting images were shuffled and separated with a 60/20/20 train/dev/test split



We used transfer leaning on the darknet implementation of YOLO. The darknet model was pre-trained on imageNet. We modified the codebase and architecture of the darknet YOLO model to account for our 15 classes rather than the original 80 classes. We re-configured our images to 718x718 and trained with batch size of 64 and subdivision size of 8 on a network with 30 layers. Our model's loss function (Standard YOLO loss function of sum-squared error between predictions and the ground-truth) improved drastically in the first 200 iterations however after this the loss only gradually decreased for the rest of training. We trained our model for around 7000 iterations before stopping training.



Input Image

Model

Model

Badly

Ground Truth



Our Model's Output



Discussion and Results

mAP = 13.36%

- We trained two YOLOv2 models, a "tiny" 15 layer model and a "regular" 30 layer model
- The larger model performed significantly better than the tiny model in class prediction and had a much lower loss.
- Our model does incredibly well at predicting tennis courts, most likely because tennis courts are large uniform objects which are quite distinct in color to their surroundings
- Our model does poorly on classes such as boats and planes because these are small objects clustered closely together. This occurs because YOLO has strong spatial constraints on bounding box predictions
- We also performed poorly on classes which were under-represented in our dataset such as
- roundabouts and helicopters
 Our model continued to improve in performance, albeit slowly, after we stopped at 7000 iterations

Conclusion / Future Work

- YOLO is versatile, but not optimal for object
- detection in satellite imagery
 Training for longer should improve performance on this task
- Data augmentation to balance dataset should improve performance on uncommon classes

References

[1] "How Many Satellites Are Orbiting the Earth in 2018?" Pixalytics Ltd, 22 Aug. 2018,

- 2016; FIASY(ILC SLU), 22 AUG. 2016, www.pixalytics.com/sats-orbiting-the-earth-2018/ [2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016
- [3] "A Large-Scale Dataset for Object Detection in Aerial Images." DOTA, captain-whu.github.io/DOTA/dataset.html.







