Generative Adversarial Networks for Satellite Image Feature Classification

Customer: Tim Parker

Subject-Matter Experts: Jonathan Brant, David Cabelly, Mark Pritt, and John Luster

Anthony DiOrio, Shane Gacke, and Milad Alaeian

Background

- **Generative Adversarial Networks (GANs)** are efficient sources for producing datasets.
- Satellite imagery present a unique problem as they are **in high demand but hard to acquire.**
- Satellite imagery cover large areas and require **extremely granular spatial resolution to facilitate object detection.**
- While there are many different computer vision methods used for identifying objects **most are not trained on satellite imagery.**
- A survey paper was created to evaluate the potential performance of different computer vision models **for a GAN that produces satellite image datasets.**

The recommended computer vision method will be able to identify 80 different objects in satellite images.

Objectives

- In order to infer the effectiveness of the computer vision methods, they will be evaluated on different criteria.
- The survey will look at different computer vision methods commonly used for object identification.
- All the methods will be evaluated in the case of the **COCO dataset**, so there is some commonality between measurements.
- After evaluating the measurements, inferences will need to be made about how they would perform for satellite images.
- Once conclusions have been made, a method will be recommended for implementation in the GAN.

Methods

- The method required for this problem must have competent **speed**, **computational requirements**, and **accuracy** with geospatial imagery.

- <u>**Speed**</u> which takes into account:
  - Graphics processing unit (GPU) timing which is how long it takes to process a single frame of an image.
  - Frames per second (FPS) on high and low-resolution images. This is related to how many images can be classified in a second.
- <u>**Computational resources**</u> which looks into:
  - The number of prediction boxes used in regards to the COCO dataset.
  - Required size for processing the image.
  - The GPU used.
- <u>**Accuracy**</u> of classification:
  - The mean average precision (mAP) of the method on small, medium and large objects.
  - The mAP for the intersection of union (IoU).

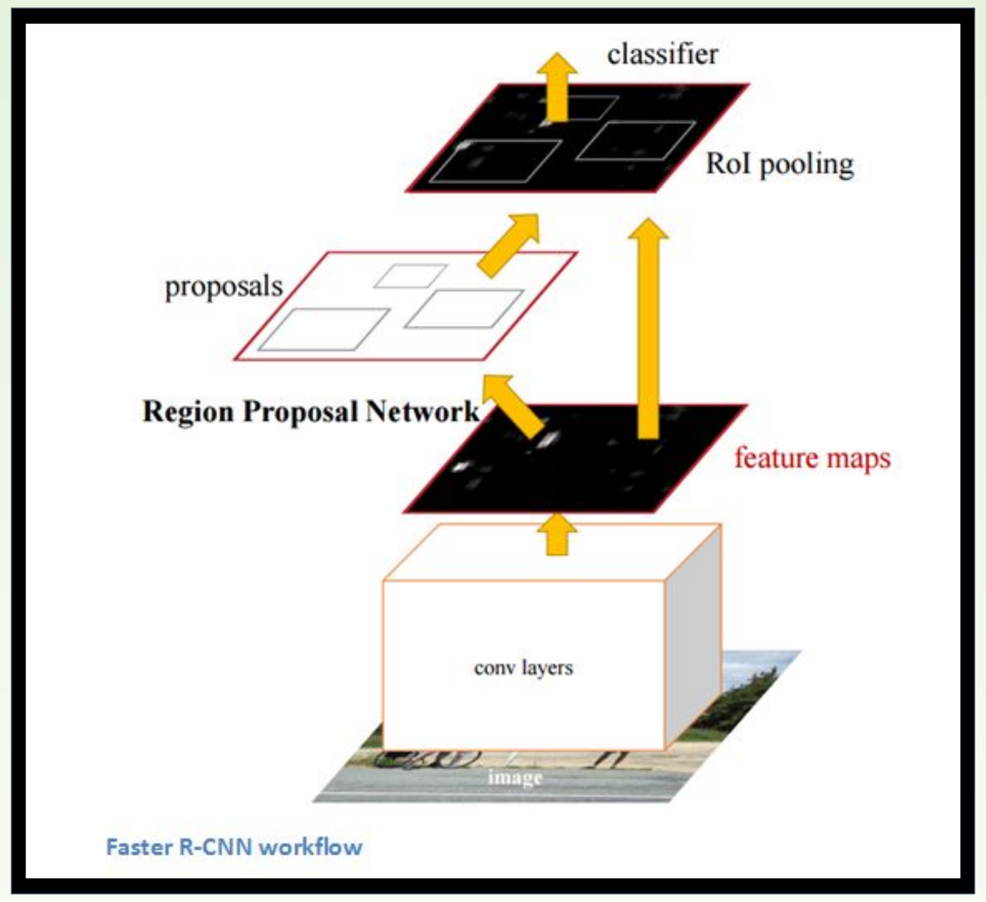
- By looking at these metrics, a conclusion can be drawn for how the method would perform with the satellite data the GAN will have to produce.

Category	Aspect	Classifier			
		Faster R-CNN	Mask R-CNN	SSD	YOLO
Speed					
	GPU Timing per Frame (milliseconds)	198 (VGG) 59 (ZF)	210 (ResNet 101)	30-170	35
	Speed High Res (FPS)	5	5	22	40
	Speed Low Res (FPS)	17	-	59	91
		Information on VGG architecture can be found at [1] and about ZF at [2]	Information on ResNet 101 can be found at [3].		
Computational Requirements					
	Number of Predictions/Boxes (COCO)	300	100*	8732	98
	Image Size Needed	Natural Image Size	Natural Image Size	300x300	448x448
	GPU Used	Nvidia Tesla K40	Nvidia Tesla M40	Nvidia Titan X	Not Listed (Assumed to be Nvidia)
			*Applied to top 100 images after processing through a region proposal network		
Accuracy (on COCO)					
	Small Object (mAP)	15.6	12.1	10.2	5
	Medium Object (mAP)	38.7	39.9	34.5	22.4
	Large Object (mAP)	50.9	52.4	49.8	35.5
	Coco IoU .5 (mAP)	48.4	62.3	48.5	44

This data from [4], [5], [6], and [7] demonstrates the capabilities of the researched vision methods.

Results

- Four different methods were found to be suitable candidates:
  - Single Shot Detection (SSD)
  - Faster Regional Convolutional Neural Networks (R-CNN)
  - Mask R-CNN
  - You Only Look Once (YOLO)
- SSD and YOLO are the most efficient in terms of speed and technical overhead.
  - This is useful for real time detection on things such as video streams.
- Mask R-CNN and Faster R-CNN perform better in terms of accuracy and adaptability to different image formats.
  - These methods are much more accurate on smaller objects in relation to the entire image
- Due to their accuracy with smaller objects, Mask R-CNN and Faster R-CNN were chosen as final candidates.
- Mask R-CNN has pixel perfect accuracy, but requires a substantial amount of data to train on.
  - Stacked GANs combined with Mask R-CNN could lead to near perfect accuracy on high resolution imagery.
- Faster R-CNN provides only bounding boxes, but does not need as much data to work on.
- Due to its greater speed and ability to work with less data, Faster R-CNN was found to be the best candidate.



A visual depiction of how the faster R-CNN method works.

Conclusions

- <u>**Final Deliverable:**</u>
  - A technical survey paper outlining research of the problem area and recommended next steps.
- <u>**Methods Researched:**</u>
  - YOLO being used to classify video in real time detection.
  - The SSD is perfect for detecting more prominent features like deforestation, while remaining relatively lightweight.
  - The Mask RCNN is the most accurate of the options.
  - Faster RCNN provides a perfect balance between speed and accuracy.
- <u>**Recommendation:**</u>
  - Faster R-CNN is the best method due to its ability to remain extremely accurate, while needing a relatively small dataset and computational requirements.

References

[1] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556v6 [cs.CV], Apr. 2015. Available: https://arxiv.org/abs/1409.1556

[2] M. D. Zeiler, R. Fergus, "Visualizing and Understanding Convolutional Networks," arXiv:1311.2901v3 [cs.CV], Nov. 2013. Available: https://arxiv.org/abs/1311.2901

[3] P. Jay, "Understanding and Implementing Architectures of ResNet and ResNeXt for state-of-the-art Image Classification: From Microsoft to Facebook," [Online]. Available: https://medium.com/@14prakash/understanding-and-implementing-architectures-of-resnet-and-resnext-for-state-of-the-art-image-cf51669e1624

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," arXiv:1506.01497v3 [cs.CV], Jan. 2016. Available: https://arxiv.org/abs/1506.01497

[5] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," arXiv:1703.06870v3 [cs.CV], Jan. 2018. Available: https://arxiv.org/abs/1703.06870

[6] W. Liu, D. Anguelove, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," arXiv:1512.02325 [cs], Dec. 2016. Available: https://arxiv.org/abs/1512.02325

[7] J. Redmon, A. Farhadi, "YOLO9000: Better, Faster, Stronger," arXiv:1612.08242v1 [cs.CV], Dec. 2016. Available: https://arxiv.org/abs/1612.08242

Acknowledgements

- We would like to thank the following people:
- Professors Sabetto and Manzo for their guidance on this project.
- Tim Parker, Jonathan Brant, David Cabelly, John Luster, and Mark Pritt for providing us with this project.
- Stephany Cando from the Hexagon capstone team for her subject matter help on the project.

May 2, 2019