Combatting Adversarial Examples in Satellite Imagery

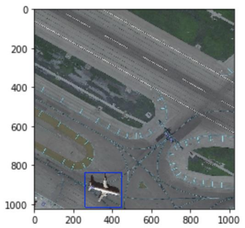
Caroline McKee (cnmckee@stanford.edu) Yash Chandramouli (yashc3@stanford.edu) Paul Caron (pcaron@stanford.edu)

Motivation

- Adversarial examples have recently been shown to successfully trick object detection networks trained on satellite imagery[1].
- Dangerous implications for national defense
- Increasingly important as systems move to real-time Idea: by using hand-selected features surrounding a detected object, can create a lightweight algorithm to boost prediction accuracy.

Data

- Data comes from (Large Dataset for Object Detection in Aerial Images)[2]
- 4000 “macro-images” with bounding box labels
- Classes: ‘ship’,’large-vehicle’,’storage-tank’,’airplane’
- Training/Val/Test: 3134/385/534 “sub-images”
- Limited data = need complex models to get insight from data



Features

Scheme	Description (size)
Sub_1	Counts of surrounding classes in sub-image (4)
Sub_2	Sub_1 + avg distances to classes in sub-image (8)
Macro_1	Counts of surrounding classes in macro-image (4)
Macro_2	Macro_1 + avg distances to classes in macro-image (8)
Macro_3	Macro_2 + avg angles to classes in macro-image (8)
Macro_4	Macro_2 and Macro_4 (12)

- Features for YOLO = CNN
- Per class for context-gen:
  - 1/(avg_dist to objects)
  - Counts
  - Average angle
- Made features for sub-images and macro-images

Models

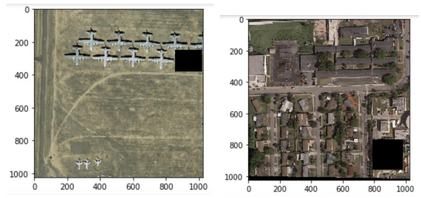
Context Algorithms

- K-NN
  - Object classified based on k nearest points
- Black Box Classifier
  - Black out object and train CNN on surrounding pixels
- Linear SVM

$$\min\left(\frac{1}{n}\sum_{\{i=1\}}^{n}\max\left(0,1-y_{\{i\}}(wx_i-b)\right)\right)+\lambda w^T w$$

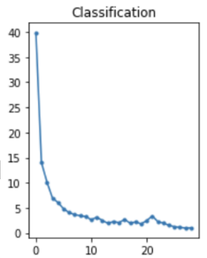
- RBF SVM: $K(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right)$
- Decision Tree
- Random Forest: Constructs multiple decision trees
- Simple NN (MLP): 3 layers, $\alpha = 1$
- Naïve Bayes: uses counts to determine probabilities
- QDA: GDA but each class has its own covariance
- AdaBoost:
  - Fitting generic weak classifiers

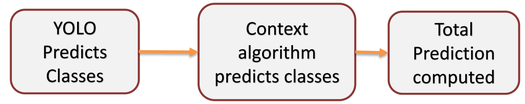
$$F(x) = sign(\sum_{m=1}^{iM}\theta_m f_m(x)),$$



YOLOv3[3]

- CNN for Object Detection and Classification
- Outputs predicted class probabilities and bounding box corners for detected objects





Discussion

- Test set and val set distributions not perfectly even, so strange results.
- Highest test-set accuracy was from Random Forest with Sub_2
- Sub-image features outperformed Macro features (not enough data, or repeated objects = overfitting)
- On average, adding features improved accuracy, as expected.
- More complex models had lower bias and higher variance (as expected)

Results

	Metric
YOLOv3	mAP = 0.449
Black Box	Acc = 0.682

	Sub_1			Sub_2			Macro_1			Macro_2			Macro_3			Macro_4		
	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test
K-NN	0.952	0.711	0.861	0.975	0.896	0.905	0.998	0.467	0.279	0.998	0.500	0.736	0.977	0.771	0.751	0.978	0.768	0.759
LinSVM	0.903	0.824	0.904	0.950	0.892	0.909	0.879	0.852	0.878	0.883	0.858	0.877	0.879	0.858	0.878	0.882	0.863	0.877
RBF	0.927	0.883	0.878	0.963	0.887	0.851	0.897	0.879	0.861	0.900	0.864	0.390	0.941	0.689	0.689	0.944	0.669	0.426
DT	0.931	0.889	0.926	0.943	0.866	0.881	0.913	0.772	0.788	0.920	0.664	0.881	0.913	0.766	0.775	0.920	0.664	0.881
RF	0.932	0.883	0.890	0.942	0.902	0.960	0.914	0.824	0.830	0.911	0.550	0.861	0.897	0.722	0.884	0.898	0.643	0.897
MLP	0.915	0.860	0.902	0.949	0.895	0.918	0.885	0.849	0.826	0.885	0.859	0.825	0.892	0.855	0.826	0.893	0.842	0.825
NB	0.844	0.819	0.840	0.879	0.882	0.895	0.811	0.886	0.762	0.763	0.767	0.598	0.802	0.883	0.780	0.768	0.781	0.594
QDA	0.846	0.820	0.838	0.882	0.884	0.870	0.809	0.885	0.762	0.770	0.755	0.591	0.811	0.886	0.768	0.783	0.780	0.603
Ada	0.918	0.897	0.869	0.930	0.884	0.804	0.849	0.818	0.783	0.867	0.229	0.867	0.862	0.864	0.887	0.875	0.227	0.871

References

[1] Czaja, W., Fendley, N., Pekala, M. J., Ratto, C., and Wang,I. Adversarial examples in remote sensing. CoRR, abs/1805.10997, 2018
[2] Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., and Zhang, L. Dota: A large-scale dataset for object detection in aerial images. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
[3] Redmon, J. and Farhadi, A. Yolov3: An incremental improvement. CoRR , abs/1804.02767, 2018
[4] github: https://github.com/yashc95/context4sats

Future Work

- Compute final boost to YOLO classification
- Potentially put context-algorithm in the loop with YOLO
- Assess time-complexity in addition