International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Dynamic Anchor Learning for Object Detection in Aerial Images

Rohit M[1], Sujay S[2], R O Sugam Kumar[3], Prof. Rekha B S[4] , Prof. Geetha V[5]

[1,2,3,4,5]*Department of Information Science and Engineering, RV College of Engineering, Bengaluru*

---***---

Abstract - *Objection detection in aerial images is comparatively a new field which is less researched. There are a few methods to detect objects in aerial images. Dynamic anchor learning method uses the recently characterized coordinating with degree to thoroughly assess the restriction capability of the anchors and completes a more productive label assignment process. In this way, the locator can progressively choose top notch anchors to accomplish precise item identification, and the dissimilarity among grouping and relapse will be lightened.*

***Key Words*: Object Detection, Computer Vision, Anchor, Ground Truth, RetinaNet, Resnet50**

1. INTRODUCTION

One of the most fundamental and difficult challenges in computer vision is object detection. Huge victories in object location have recently been achieved thanks to the improvement of convolutional neural networks (CNN). Most identification structures use preset flat anchors to accomplish spatial arrangement with ground truth(GT) boxes. Positive and negative examples are then chosen through a particular system during the preparation stage, which is called label assignment.

There are many approaches which have achieved aerial image object detection by introducing the extra orientation prediction and preset rotated anchors. In this paper, the approach named “dynamic anchor learning” is discussed.

2. LITERATURE SURVEY

This paper focuses on a new approach for extracting multi-scale strong and weak semantic features that uses a novel hybrid image cascade and feature pyramid network with multi-size convolution kernels.[1].

This study examines an innovative and successful method for learning a rotation-invariant CNN (RICNN) model for improving object identification performance, which involves introducing and learning a new rotation-invariant layer on top of current CNN architectures.[4].

This study focuses on a method for forecasting the localization uncertainty, which is used to determine the bbox's reliability. The proposed techniques can reduce the false positive (FP) and raise the true positive (TP) by exploiting projected localization uncertainty throughout the detection process, hence enhancing the accuracy.[6].

This study examines uncertainty-aware learning from demonstration technique, offering a unique uncertainty estimation method based on a mixed density network that is suitable for modelling complicated and noisy human actions.[8].

This study uses convolutional neural networks in its entirety which results in detection of objects using lesser computational resources and at an improved speed. This also has the effect of sharing the computation performed throughout the image. This is in direct contrast with other CNNs that perform region-based detections i.e, Fast/Faster R-CNN, which run a computation heavy code repeatedly on an image. [11].

3. METHODOLOGY

Python is utilized to assemble the product. Pytorch is used for developing and training the model. The dataset is used to train the model to detect the cars and aircraft's in the image. After the model is trained it will be able to detect cars and aircraft's.

Data collection

Images were used from the dataset named UCAS-AOD. It is an aerial aircraft and car detection dataset, which contains 1510 images.

Data pre-processing

The images are already annotated in the UCAS-AOD dataset, which we are using.

Model training and testing

This task is aimed at training the model in order to detect the cars and aircrafts. The chosen aspect ratio for the three horizontal anchors are {1/2,1,2}. All images are resized to 800×800. We use random flip, rotation, and HSV colour space transformation for data augmentation. The total iterations of the images in the dataset are 15k.

4. SYSTEM ARCHITECTURE

We use a model built on one stage detector RetinaNet, which has Resnet50 as it’s backbone to construct a multiscale feature pyramid. This backbone is connected to two subnetworks: one for anchor box classification and the other for regressing from anchor boxes to ground truth object boxes.

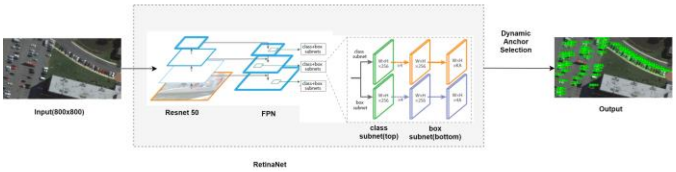

Fig -1: System Architecture

We feed the data obtained to calculate a matching degree, which inturn is used to find the matching sensitive loss. When the model is trained with the matching sensitive loss, we achieve a higher detection in objects when compared to the baseline which is just a RetinaNet.

5. PERFORMANCE ANALYSIS

We check the performance of the model by plotting a confusion matrix and calculating the precision, recall and f1-score.

Testing

A training set containing 100 images was prepared and the model was tested on it.

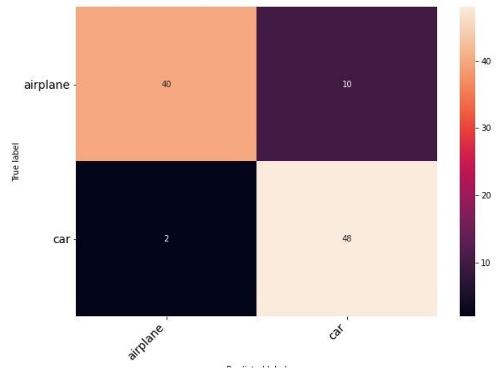

Fig -2: Confusion Matrix

X-axis - Predicted Values

Y-axis - Actual Values

airplane - 40/50 detected accurately

car - 48/50 detected accurately

Table -1: Classification Report

	precision	recall	f1-score	support
airplane	0.95	0.80	0.87	50
car	0.83	0.96	0.89	50
accuracy			0.88	100

Number of images used in the dataset:

Airplane - 50 images

Car - 50 images

So, all the images total to 100 images. Of which the precision of airplanes detected is 95% and the precision of cars detected is 83%.

6. CONCLUSION

Dynamic anchor learning method helps in efficiently detecting cars and aircraft's in aerial images. Currently there are very few methods to detect objects in aerial images, where only the top view of the images are visible. Based on the results obtained it can be said that this model is efficient to detect objects in aerial images.

REFERENCES

[1] Azimi, S. M.; Vig, E.; Bahmanyar, R.; Korner, M.; and ¨ Reinartz, P. 2018. Towards multi-class object detection in unconstrained remote sensing imagery. In Asian Conference on Computer Vision, 150–165. Springer.

[2] Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; and Lu, Q. 2019. Learning RoI transformer for oriented object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2849–2858.

[3] Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. International journal of computer vision 88(2): 303–338.

[4] Cheng, G.; Zhou, P.; and Han, J. 2016. Learning rotation invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 54(12): 7405–7415.

[5] Feng, D.; Rosenbaum, L.; and Dietmayer, K. 2018. Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 3266–3273. IEEE.

[6] Choi, J.; Chun, D.; Kim, H.; and Lee, H.-J. 2019. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, 502–511.

[7] Han, J.; Ding, J.; Li, J.; and Xia, G.-S. 2020. Align Deep Features for Oriented Object Detection. arXiv preprint arXiv:2008.09397.

[8] Choi, S.; Lee, K.; Lim, S.; and Oh, S. 2018. Uncertainty aware learning from demonstration using mixture density networks with sampling-free variance modeling.

In 2018 IEEE International Conference on Robotics and Automation (ICRA), 6915–6922. IEEE.

[9] Jiang, B.; Luo, R.; Mao, J.; Xiao, T.; and Jiang, Y. 2018. Acquisition of localization confidence for accurate object detection. In Proceedings of the European Conference on Computer Vision (ECCV), 784–799.

[10] Jiang, Y.; Zhu, X.; Wang, X.; Yang, S.; Li, W.; Wang, H.; Fu, P.; and Luo, Z. 2017. R2cnn: rotational region cnn for orientation robust scene text detection. arXiv preprint arXiv:1706.09579.

[11] Dai, J.; Li, Y.; He, K.; and Sun, J. 2016. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, 379–387.