International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Object Classification Using Satellite Images

A Dharma Reddy[1], P Satya Sriram[2], R Shreya[3], Riya Naresh[4], Dr. Anala M R[5] ,Prof. Ganashree KC[6].

[1-4]Student, Department of Computer Science & Engineering, R V College of Engineering, Karnataka, India
[5]Professor, Department of Computer Science, RV College of Engineering, Bangalore, India.
[6]Associate Professor, Department of Computer Science, RV College of Engineering, Bangalore, India.

---***---

Abstract - *Satellite images are so important in numerous applications which include natural calamities management, auto-pilot feature of self-driving cars, law implementation and contempt, satellite imaging and many more. For all these applications, we require crucial attention from humans in observing changes in the environment. As we know that the area on the surface of earth is too vast to be easily covered manually and also poses high difficulty in monitoring it, we have to shift to automation to use the limited resources that are available to us efficiently. Since the existing old methods and algorithms are not very efficient to resolve the issues. Deep learning is a family of machine learning algorithms that have shown significant promise for the automation of such tasks. The existing techniques in computer vision and object recognition depend vigorously on CNN and DNN. The object detection model draws a bounding box around the object in rectangular, horizontal and vertical shapes to show the position of the object in the satellite image. This model developed to detect multiple objects in the image. It is brought into life in Python language using the Keras and TensorFlow deep learning libraries. These models can be used in autonomous drone detection and tracking systems. The DOTA dataset is used in the development of this model. The model is trained with 10000+ images and tested on 400+ images*

Key Words: Deep Learning; Machine Learning; Convolutional Neural Network; Classification; Satellite Imagery.

1.INTRODUCTION

Computer vision has taken incredible steps in the course of recent times. It is currently a necessary part of everybody's lives as it is utilized all over the place, from medical equipment to modern assembling. Numerous undertakings, for example, independent and autonomous driving, facial acknowledgment and optical character acknowledgment would not be conceivable without computer vision. The significant level of goal of Computer Vision is to separate data from high dimensional, certifiable pictures from which to decide, for example spotting a traffic signal becoming red and apply brakes to stop the vehicle. The way toward extricating such data from a picture can shift. Over the previous decade, profound learning strategies utilizing convolutional channels has arisen to be the main procedure for Computer vision, helping assignments like picture rebuilding, object acknowledgment, present situation assessment and movement assessment etc.

In the field of Computer vision, two undertakings overwhelm research, they are: object detection and object classification. Performance depends on how accurately perceiving or characterizing an image of an object is done. In profound learning, this is finished utilizing a neural organization with c number of outputs where c is the quantity of objects the organization has been prepared on. Object identification isn't just perceiving the article, yet additionally precisely limiting it inside a picture. Therefore, object recognition is essentially more perplexing and is a space of extraordinary examination. Profound learning utilizes Convolutional Neural Networks (CNN's) to remove data from high dimensional images. The paths required for such convolutional activities are learned by the network through backpropagation. Such path weights are known as boundaries or loads. They may additionally be joined by a predisposition or a bias that moves the mean of the operation from beginning to some other point on a cartesian plane. A regular CNN can have anyplace from a little modest bunch to over 100 convolution tasks. The convolutions are normally mediated with nonlinear actuations and standardization operations. Examination of object recognition however isn't restricted to improving general discovery correctness or improving discoveries for explicit classes of items, but also to explicitly measured objects, improving the speed of recognition or diminishing computing power and memory footprint.

YOLO - You Just Look Once [15] is a profound neural network technique utilized for object recognition. YOLO eliminated such multi-stage preparation, lessening learnable boundaries and network intricacy. By smoothing out this cycle and lessening of learnable parameters, training and inference was made quicker in YOLO. In YOLO, non-max suppression is done avariciously. The bounding box with the most certainty is picked. All bounding boxes with high IOU with this chosen bounding box are disposed of. Of the excess bouncing boxes, the one with the most elevated certainty is chosen next and the cycle proceeds. Notwithstanding being a best in class object identification network with higher than constant induction times, YOLO has its inadequacies. Particularly in examinations with different networks that were on top of the leader board at that point, YOLO gave lower mAP scores. Like the base YOLO network, YOLO v3 likewise passes the image just a single time through the network prior to making a forecast, subsequently holding the "only looking once" highlight of all YOLO networks. This is one of the key highlights which add to its continuous nature. The architecture can be partitioned into 3 sections - the backbone, the prediction feature maps, and the loss. This is worthwhile, since it permits free testing and adjustment of any each module in turn. Darknet-54 has 54 convolution layers that resize an image from W H to prediction maps of three different scales. Native YOLO v3 falls short in providing tightness of bounding boxes when objects are placed orthogonally. This is detrimental to computer vision, so it uses oriented bounding boxes.

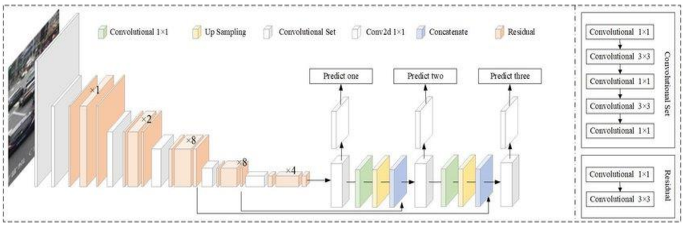

Figure 1. YOLOv3 Architecture

DOTA [1] is used here and has a huge scope aeronautical imaging dataset that was caught by satellites. It comprises 1411 pictures of different goals and angle proportions.

2. Related Work

A few explained datasets of symbolism, alongside identification and classification, have shown up. A large portion of the profound learning applied to distantly detected symbolism has managed land cover arrangement or building identification. For an instance, the University of California Merced Dataset consisted of twenty one hundred aerial pictures from the United States Geographical Department [4,5]. These Images consisted of 256x256 pixels of size with a ground test distance of 0.3 mts per pixel of image . It has 21 classes which have details about land coverage types like agro lands, roads, technology building classes and water bodies, for example, stockpiling tanks and football courts. A few developers utilized the Residual Neural Network, and Inception CNNs to order the University of California Merced pictures into land surface types [9-10] and one announced grouping correctness's as high as 98.5% [10]. The current dataset is exceptionally restricted, notwithstanding, in its dimensions, the number and kinds of classes, and geographic variety.

The SpaceNet dataset [11] comprises high level Digital Globe satellite pictures of four urban areas alongside with building CNNs that have been prepared to section the pictures and concentrate on buildings [12]. This dataset has limitations regarding its various geographical surface coverage and efficiency in modelling a classifier.

Reference to [13], it contains a rundown of many differing detecting datasets. But not even a single dataset consists of countless pictures on a worldwide scale that are needed to foster a flexible picture order framework.

One of the main profound networks to offer object recognition was R-CNN [18] by Girshik et al. It utilized specific pursuit techniques to initially extricate parameters from a picture and afterward passed every locale through an article acknowledgment Convolution Neural Network (CNN). This was not just extremely sluggish but on the other hand was extremely subject to the particular hunt calculation.

Vital networks to carry out object location, was You Only Look Once - YOLO [16]. YOLO at first tended to speed as its essential commitment while giving practically identical discovery scores. At the time it was one of only a handful of network models to offer ongoing identification and classification. Resulting upgrades to the YOLO network included YOLO v2 [17] and YOLO v3 [19]. These are explained in higher detail in the later areas of the archive.

3. Dataset

The images for this project are taken from DOTA which are collected using Google Earth and GF2 and JL1 satellites supported by China for the research in the field of Satellite object detection . DOTA is a huge scope flying imaging dataset that was caught by satellites. It comprises 1411 pictures of different resolutions and aspect ratio. The aspect ratios change from 1024 x 1024 to as outrageous as 500 x 3000. The dataset comprises object categories: soccer ball field, storage tank, baseball diamond, ground track field, harbor, plane, bridge, large vehicle, small vehicle, helicopter, ship, tennis court, swimming pool, basketball court, roundabout and container crane.

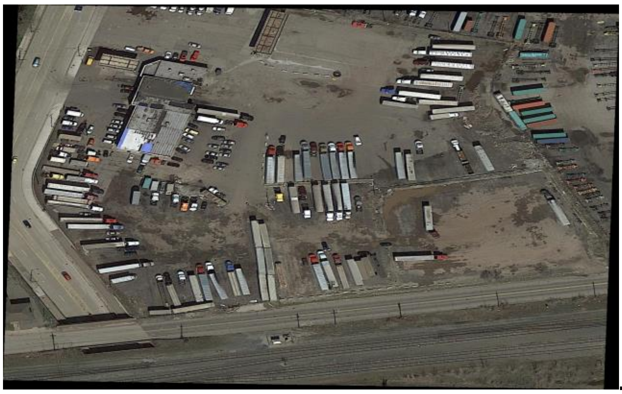

Fig 2: Sample Image in DOTA dataset

Fig 3: Sample Image in DOTA dataset

In the DOTA dataset, the objects are annotated by OBB . The bounding box vertices are numbered in clockwise direction. The objects inside the bounding box are labelled with a class name from among the 15 classes. A difficulty is mentioned, which lets us know the level of difficulty for an object to be detected and classified (for difficult - 1, for not difficult - 0). Image annotations are then saved in a text file having the same name.

4. Proposed Methodology

The proposed method detects and classifies the various objects in the satellite image.

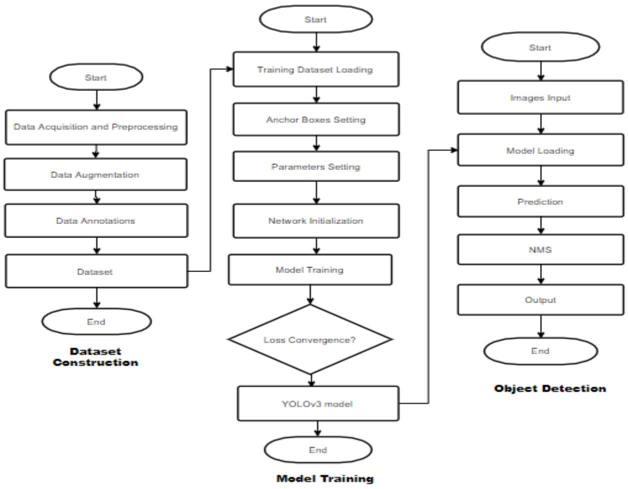

Fig 4: Flowchart of model

4.1. Image processing

The satellite pictures are of high resolution and the size of the pictures are high. Since the YOLO model is skeptic to picture size, the actual design doesn't go against the wide varieties in aspect ratio or picture size. But, utilizing such pictures diminishes the overall size of objects inside the picture, making it harder for the model to learn. Consequently, the pictures are trimmed to 1024 x 1024 lumps from huge and uneven sizes of 3000 x 1000. This is finished by utilizing a sliding window approach [3] . Cropping guarantees that the pictures are all of equivalent sizes and aspect ratio.

4.2. Model Network

Darknet-54 has 54 convolution layers that resize a picture from W x H to forecast guides of three distinct scales [2]. The image goes through different convolutions, each utilizing leaky ReLU and batch normalization. It ought to be noticed that there are skip associations between layers to avoid vanishing gradients. Skip connections are added between layers by an additional operation. Due to this all the three dimensions do not change, only the value at each pixel changes. It ought to be noticed that instead of pooling, the picture is resized utilizing convolution activities with a step of 2. Two convolutional layers with a skip layer contribution from a former layer will be alluded to as a resblock.

Layer	Filters size	Repeat	Output size
Image			416 × 416
Conv	32 3 × 3/1	1	416 × 416
Conv	64 3 × 3/2	1	208 × 208
Conv Conv Residual	32 1 × 1/1 64 3 × 3/1	Conv Conv Residual × 1	208 × 208 208 × 208 208 × 208
Conv	128 3 × 3/2	1	104 × 104
Conv Conv Residual	64 1 × 1/1 128 3 × 3/1	Conv Conv Residual × 2	104 × 104 104 × 104 104 × 104
Conv	256 3 × 3/2	1	52 × 52
Conv Conv Residual	128 1 × 1/1 256 3 × 3/1	Conv Conv Residual × 8	52 × 52 52 × 52 52 × 52
Conv	512 3 × 3/2	1	26 × 26
Conv Conv Residual	256 1 × 1/1 512 3 × 3/1	Conv Conv Residual × 8	26 × 26 26 × 26 26 × 26
Conv	1024 3 × 3/2	1	13 × 13
Conv Conv Residual	512 1 × 1/1 1024 3 × 3/1	Conv Conv Residual × 4	13 × 13 13 × 13 13 × 13

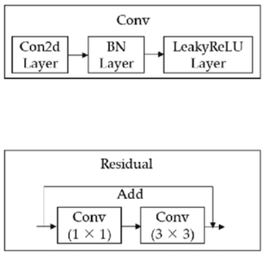

Fig 5: Layers in DarkNet - 53 [16]

4.3. Training

Transfer Learning: This is a kind of technique in which the gained knowledge by a machine is applied to a similar problem so as to reduce the training computation. In this project, we have used a pre-trained model which is trained using images in ImageNet dataset. This trained model is thus taken and fine-tuned using images in DOTA dataset. We have found some pre-trained models that were available online and included in Keras [7,8] whereas few were taken from the deep learning community. Although pre-trained models of ImageNet dataset do not contain satellite images, it has shown efficient transfer learning features towards other domains, here satellite images [9,14]. So we have used this model to begin our training with DOTA dataset, this has reduced our training period as well as given an increased accuracy rate for classification of objects in satellite images.

Data Augmentation: Using a sliding window approach the images are splitted and resized to 1024 x 1024. And the image number is increased due to this technique.

Epochs: We have a 2 stage training where the layers are freezed and trained to get a stable loss and then the layers are unfreezed and trained for better results. So, in the first stage we used 25 epochs and second stage as more GPU memory is used when unfreezed we have used 10 epochs.

Learning Rate: Lr used is Binary cross entropy. Using some keras call backs the learning rate is reduced if there is no significant improvement in the model.

In the first stage the learning rate is 0.001 and in the second stage it is 0.0001.

Hyper-parameters used:

- Number of Epochs - 35
- Loss Function - cross entropy
- Optimizer - Adam
- Learning Rate - 0.001, 0.0001

5. Experimental Results

The dataset has 1800+ images for training with labels which are splitted and cropped to increase the dataset and balance the aspect ratio. The model is trained with 10000+ images.

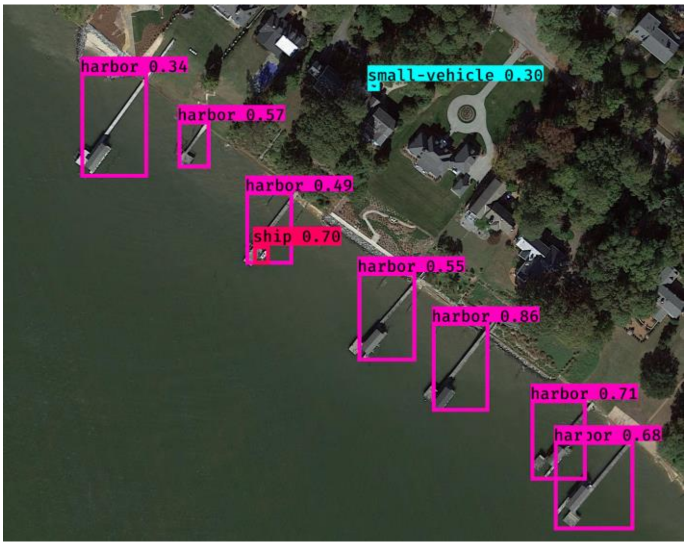

Fig 6: Harbors detected in the image

Fig 7: Planes detected in the image

The model detects more than 80% of the objects with an accuracy of 75 and more for the images provided.

6. Limitations

There is a problem with several competing systems on the marketplace using arrays of technologies. These competing systems may offer similar functions or capabilities through alternative technical approaches. We are able to classify all the class objects in the test satellite images. The model is efficient enough to classify all classes of the objects but not with 100 percent accuracy and can be improved with more dataset. The hardware support has to be intensified. The bounding boxes are not rotated according to the object.

7. CONCLUSIONS

This project has successfully demonstrated the capability of a deep learning system that can efficiently classify objects which were present in an high-resolution satellite image which offers multi-spectral features. The proposed system uses a yolo model predicting the objects within a bounded box.

We believe that this could be very beneficial to areas of surveillance, defence and tracking. Despite this, we hope that the work presented here is used for the benefit of humankind and is used to improve quality of life and progress research in deep learning. We aim to improve the model to detect more objects with better efficiency.

ACKNOWLEDGEMENT

We would like to express our deep gratitude towards our Guide Associate Prof. Dr Anala M.R and Prof Ganashree KC, for providing an innovative project idea and their deterministic guidance, humble nature and experienced suggestions have helped a lot in successfully completing this case study paper. Also we whole-heartedly thank our beloved Principal Dr. K. N. Subramanya for their continuous support to carry out this project.

REFERENCES

[1] D. Ni and Z. Chang, "DOTA: A Large-scale Dataset for Object Detection in Aerial Images ," 2018. [Online]. Available : https://captain-whu.github.io/DOTA/index.html

[2] Joseph Redmon, Ali Farhadi "YOLOv3: An Incremental Improvement"Available:https://pjreddie.com/media/files/papers/YOLOv3.pdf

[3] DOTdevelopment kit :https://github.com/CAPTAIN-WHU/DOTA_devkit

[4] "UC Merced Land Use Dataset," University of California, Merced,http://weegee.vision.ucmerced.edu/datasets/landuse.html.

[5] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," Proc. 18th ACM SIGSPATIAL International Symposium on Advances in Geo. Info. Sys., pp. 270-279, 3-5 Nov 2010.

[6] Structure of darknet https://www.researchgate.net/figure/Structure-of-the-Darknet53-convolutional-network_fig4_338121987

[7] "Applications", Keras, https://keras.io/applications/.

[8] https://github.com/flyyufelix/Cnn_finetune.

[9] Y. Liang, S. Monteiro, and E. Saber, "Transfer learning for high-resolution aerial image classification," IEEE Workshop AppliedImagery Pattern Recognition (AIPR), Oct 2016.

[10] G. Scott, M. England, W. Starms, R. Marcum, and C. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery", IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 4, pp. 549-553, Apr 2017.

[11] "SpaceNet on AWS," Amazon.com, https://aws.amazon.com/public datasets/spacenet/.

[12] E. Chartock, W. LaRow, and V. Singh, "Extraction of Building Footprints from Satellite Imagery," Stanford University Report, 2017.

[13] G. Cheng, J. Han, and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," Proc. IEEE, vol. 105, no. 10, pp. 1865-1883, October 2017.

[14] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" Proc. 31st Int. Conf. Machine Learning, vol. 32, pp. 647-655, 2014.

[15] J. Redmon, S. Divvala, R. Girshik and F. Ali, "You Only Look Once: Unified, Real-Time Object Detection," in Computer Vision and Pattern Recognition, 2015.

[16] 1] J. Redmon, S. Divvala, R. Girshik and F. Ali, "You Only Look Once: Unified, Real-Time Object Detection," in Computer Vision and Pattern Recognition, 2015.

[17] J. Redmon and A. Farhadi, " YOLO9000: better, faster, stronger," in Computer Vision and Pattern Recognition, 2017.

[18] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 22 October 2014. [Online]. Available: arXiv:1311.2524 [cs.CV].

[19] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 8 April 2018. [Online]. Available: arXiv:1804.02767 [cs.CV]