JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Object Detection Using Detectron

P. K.V. Subbaraya Sarma[1], B. Vishal Reddy[2], V. Tarun Kumar[3], Y. Umamaheshwar[4], Y. Sathya Manoj Ram[5]

[1]Assistant Professor, Department of Computer Science and Engineering, Vidya Jyothi Institute of Technology, Aziz Nagar, Hyderabad, Telangana, India

[2,3,4,5]Student, Department of Computer Science and Engineering, Vidya Jyothi Institute of Technology, Aziz Nagar, Hyderabad, Telangana, India

***Abstract*—**The thing of this exploration is to ameliorate object discovery using the detectron2 codebase, which was developed by the Facebook AI exploration (FAIR) platoon. The identification and localization of particulars within an image or a videotape are done using the object discovery fashion in computer vision. By spatially segregating bounding boxes and employing a single convolutional neural network to assign changes to each of the detected images, the generators of the YOLO (You Only Look formerly) fashion framed the object identification problem as a retrogression problem rather than a bracket task (CNN). As we can easily see, this system has some downsides. For illustration, it has lower recall and lesser localization error than Faster RCNN, struggles to identify near objects because each grid can only suggest two bounding boxes, and has trouble relating small objects. In discrepancy, if we choose styles grounded on region proffers, quicker RCNN uses a model that combines a region offer network and a point aggregate network. In order to address the issues, we see as being object discovery ways, this paper uses the hastily RCNN rather than the YOLO approach.

Keywords—Faster RCNN, Object detection, Region Proposal Networks, Feature Pyramid Networks, Deep Learning, ROI pooler, Transfer Learning, Image Recognition, Computer Vision, Convolutional neural networks (CNNs), Feature extraction, Object localization, Object classification, Anchor boxes, Bounding box,

IoU (Intersection over Union), Two-stage detectors, ResNet (Residual Neural Network).

1 Introduction

Object Discovery is a computer vision fashion for locating cases of objects in images or videos. Object discovery algorithms generally work with machine literacy or deep literacy to produce meaningful results. When humans look at images or videos, we can fete and detect objects of interest within a matter of moments. The thing of object discovery is to replicate this intelligence using a computer. originally, we can train a custom object sensor from scrape and need to design a network armature to learn the features of the objects of interest. Also, need to collect a veritably large set of labeled data to train the CNN. The results of a custom object sensor can be remarkable. The CNN's layers as well as weights must be explicitly set up, which requires considerable time and training set. Another approach we can follow is that numerous object discovery workflows using deep literacy influence transfer literacy, an approach that enables you to start with a pre-trained network and also fine melodies it for your operation. This system can give faster results because the object sensors have formerly been trained on thousands, or indeed millions, of images. We can produce a custom object sensor or use a pre-trained one, you'll need to decide what type of object discovery network you want to use, a two-stage network or a single-stage network. We also have single and two-stage networks available in the object discovery algorithms. Single-stage discovery algorithms similar to YOLO will descry regions across the entire image using anchor boxes and prognostications are decrypted to induce the final boxes. Whereas the two-stage discovery algorithm similar to faster RCNN will identify regions in the first phase and classify the linked objects in the alternate phase along with the addition of the bounding boxes.

2 Literature review

Multitudinous similar publications have formerly been published and are available. This section will examine applicable check ways for issues and their fixes before extending them to produce an operation. These are some reviews of the literature.

1. Mr. K. Bharath's" Object Detection Algorithms." Both pretensions of the algorithms for object discovery and how

they work have been outlined in detail by him. It focuses on the trouble that went into the object discovery algorithms and how we may ameliorate them going forward. He also gives us an overview of the most recent libraries and software programs that we can use to descry objects in picture or videotape data.

2. The two specific object discovery phases that distinguish single- and two-phase networks in the algorithms are the main content of this paper. By describing how these networks serve outside, we have the inside track on how to ameliorate them in the future.

3. Ross Girshick, Jeff Donahue, Trevor Darell, and Jithendra Mallik have created papers named" Rich point scales for object discovery and semantic segmentation" in which they detailed the former object discovery styles and how they would serve. They gradationally move on to bandy the elaboration of object discovery algorithms and their practical operations.

4. Karen Simonyan and Andrew Zisserman's "Very Deep Convolutional Networks for Large-Scale Image Recognition," in which they investigate the effect of convolutional network depth on accuracy in the large-scale image recognition setting. Their key contribution is a detailed study of increasing depth networks utilizing an architecture with extremely small (3x3) convolution filters, demonstrating that increasing the depth to 16-19 weight layers tends to result in a substantial improvement over the previous setup. They used these results as the basis for their entry to the 2014 ImageNet Challenge, which brought them first and second place in the localization and classification categories, respectively. They also demonstrate that our representations generalize well to other datasets, achieving state-of-the-art results. To promote further study on the deep visual representations' application in computer vision, the researchers have publicly released their two best performing ConvNet models.

5. In their article "Mask R-CNN", Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick presented a conceptually straightforward, adaptable, and all-encompassing framework for object instance segmentation. Their method successfully locates items in an image while producing a top-notch segmentation mask for each instance. By introducing a branch for predicting an object mask along with the current existing branch for bounding box detection, their technique, Mask R-CNN, expands Faster R-CNN. Faster R-CNN is performed at five frames per second while Mask R-CNN creates only a small overhead. Additionally, within the same context, Mask R-CNN is readily adaptable to other tasks like estimating human poses. Instance segmentation, bounding-box object detection, and person key point detection—three COCO challenge tracks—they achieved outstanding outcomes in each. On every job, including the COCO 2016 challenge winners, Mask R-CNN performs better than any single-model entry currently in use.

6. In their article titled "Faster R-CNN," Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun describe a state-of-the-art object detection network that relies on region proposal algorithms to make location predictions for objects. The running time of these detection networks has decreased thanks to innovations like SPPnet and Fast R-CNN, revealing region proposal computation as an obstruction. In this work, they present a Region Proposal Network (RPN) that collaborates with the detection network to share full-image convolutional features, allowing almost cost-free region proposals. A pure convolutional network known as an RPN forecasts object bounds as well as object scores at each location at the same time. The RPN is thoroughly taught to produce superior region proposals, which Fast R-CNN uses for detection. By combining the convolutional features of RPN and Fast R-CNN, they were able to further combine the two networks into one. The RPN component instructs the combined network where to look, in line with the newly well-known concepts of neural networks using "attention" mechanisms. On the PASCAL VOC 2007, 2012, and MS COCO datasets, our detection system on a GPU achieves state-of-the-art object recognition accuracy with only 300 proposals per image at a frame rate of five frames per second (comprising all steps) for the very deep VGG-16 model. Faster R-CNN and RPN are the cornerstones of the first-place winning entries in a number of tracks in the ILSVRC and COCO 2015 contests.

3 Existing Methodology

Item detection in one-stage Models is a class of one-stage object detection algorithms, which bypass the region proposal stage of 2 stage models and perform detection immediately over a large number of locations. These models typically have quicker inference (possibly at the cost of performance). On the other hand, a one-stage detector simply needs one pass through the neural network and calculates all of the bounding boxes at once. It is significantly quicker and more suited for mobile devices. One-stage object detectors like YOLO, SSD, SqueezeDet, and DetectNet are the most popular types. As I noted before, because of the nature of our "predictions on a grid" approach, we frequently wind up with a significant number of bounding boxes in which no object is enclosed. After establishing a defined set of bounding box predictions, we can simply filter these boxes out, yet there remains a (foreground-background) class imbalance that can cause issues during training. This is particularly challenging for models that merely have included a "background" class for regions without any objects instead of splitting the prediction of objectness and class likelihood into two distinct tasks. Researchers at Facebook suggested modifying the standard cross entropy

loss by including a scaling factor to emphasize "challenging" cases more during training and prevent the dominance of simple negative predictions.

4 Proposed Methodology

This study illustrates the process of object detection utilising the faster RCNN state-of-the-art algorithm, which was implemented with the aid of detectron2, a codebase created by the Facebook AI Research (FAIR) team to facilitate the implementation of such cutting-edge algorithms.

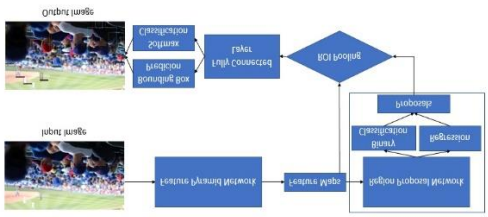

Fig 1: Architecture of Proposed System

Concept and design: At the moment, there are two different types of networks: single-stage and two-stage. In single-stage networks, there is only one process that uses anchor boxes to predict regions across the entire image, and the predictions are then decoded to produce the final bounding boxes for the objects. A region proposal, or a subset of the image that may include an item, is identified in the first stage of two-stage networks like R-CNN and its variations. The objects contained in the region proposals are categorised in the second step. Single stage networks like YOLO have an extremely short processing time, whereas two-staged networks are great for accurate identification.

4.1 Component in play

4.1.1 The video or picture will be uploaded by the user to the interface and processed from there.

4.1.2 System: It will accept input from the user and carry out two processes simultaneously, namely the extraction of picture features using Feature Pyramid Networks (FPN) and the detection of likely object-containing regions using region proposal networks (RPNs). These will then be sent to a classifier, such as a support vector machine, for the classification of the identified items after being transmitted to the ROI (Region of Interest) pooler.

4.2 Implementation & Method: The paper explains the steps taken in the quicker RCNN object identification. We have demonstrated the operation of this procedure with thorough UML diagram illustrations.

The improved version of Fast RCNN is called Faster RCNN. Faster RCNN uses a "Region Proposal Network," also known as RPN, whereas Fast RCNN uses selective search to generate Regions of Interest. RPN creates a series of object proposals with an objectness score for each one after receiving input from image feature maps.

The following actions happen in the Faster RCNN: The ConvNet receives an image as input and produces the feature map for that picture. These feature maps have a region proposal network added to them. The object proposals and their objectness score are returned. Then, to make all of the proposals the same size, a RoI pooling layer is applied to these proposals. In order to categorize and output the bounding boxes for objects, the proposals are finally given to a fully connected layer that has a softmax layer and a linear regression layer at its top.

To identify the objects, every object detection algorithm we've studied so far makes use of regions. The network focuses on different areas of the image in turn rather than viewing the entire thing at once.

Two problems result from this:

• The algorithm needs to go over a single image multiple times in order to extract all the objects.

• Because several systems operate sequentially, the effectiveness of the systems that follow is influenced by the effectiveness of the ones that came before.

5 Results and discussions

Below are the steps which help us to use the web application hosted with the help of the Streamlit package.

Step-1: The web application opens with the page which shows its title, the “about the app” drop-down menu, and an option to upload the image.

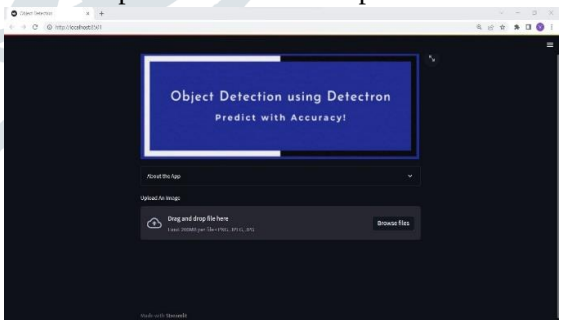

Fig 6: Landing Page

Step-2: To know some information about the application, click on the drop-down icon to view the given information.

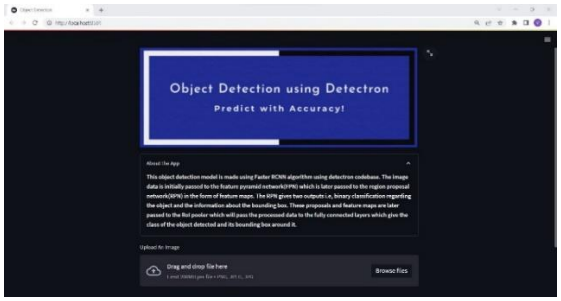

Fig 7: Markdown Page

Step-3: Click on browse files to upload an image to pass it to the network for detecting the objects present in it.

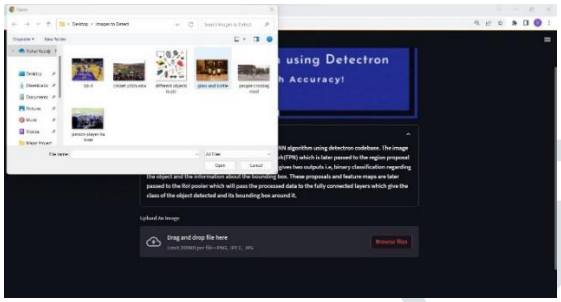

Fig 8: Uploading Image

Step-4: The image will be uploaded on the web application, and the "File Saved" message will be displayed at the bottom of the screen.

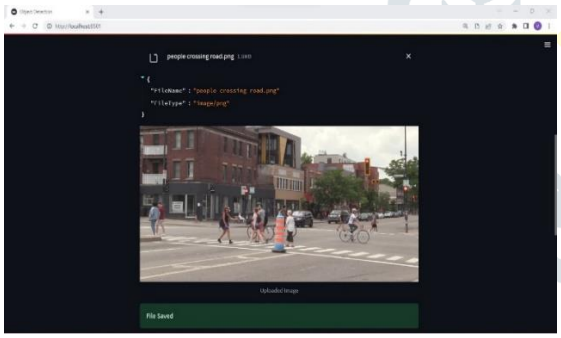

Fig 9: Input Image

Step-5: After a few moments, we can see the bounding boxes around the image along with the class of the image and the confidence rate.

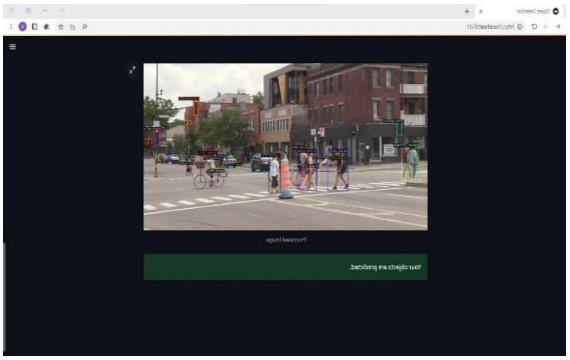

Fig 10: Output Image

Step-6: We also have the option to enlarge the image and review it.

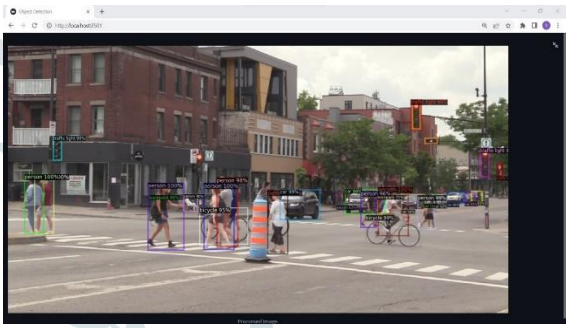

Fig 11: Enlarged View of Image

6 Conclusion

With the help of sophisticated object tracking as well as video segmentation software known as Detectron2, we can build cutting-edge algorithms such as fast RCNN, faster RCNN, and others that improve performance and speed in real time. The Intersection of Union (IoU) value can be changed to correct this issue with faster RCNNs when the bounding boxes occasionally overlap the objects that are detected surrounding the neighbouring image. We demonstrated that quicker RCNN using detectron2 performs effectively with picture data since it is a cutting-edge approach made up of clustering different architectures and techniques. We found that while single-stage networks are quicker, faster RCNN detects more significant items than other single-stage algorithms. Faster RCNN has consistently produced good results based on the input data when it comes to tightened bounding boxes. Also, it was able to find tiny items in images that were mingled with large crowds. Overall, the faster RCNN with the detectron2 codebase performs flawlessly with the input data.

7 References

[1] Ren, S.Q., He, K.M., Girshick, R., Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. Montreal.2016, pp. 91-99.

[2] Convolutional Neural Networks (CNNs): http://cs231n.github.io/apr 2020

[3] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, p. 1627, 2010.

[4] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data." IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–15, 2017.

[5] Ren, S.Q., He, K.M., Girshick, R., Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. Montreal.2016, pp. 91-99.

[6] J. Redmon, S.Divvala, R.Girshick, A. Farhadi, You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788)(2016). https://doi.org/10.1109/cvpr.2016.91.

[7] Torralba, A., Fergus, R., Freeman, W.T. 80 million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence,2008, pp.1958-1970.

[8] X. Peng, C. Schmid, Multi-region two-stream R-CNN for action detection. In European conference on computer vision (pp. 744-759). Springer, Cham. (2016, October).https://doi.org/10.1007/978-3-319-46493-0_45.

[9] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99) (2015). https://doi.org/10.1109/tpami.2016.2577031.

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

[11] C.Y. Fu, W. Liu, A.Ranga, A. Tyagi, AC Berg, Dssd: Deconvolutional single shot detector. arXiv preprint arXiv:1701.06659. (2017).

[12] Jiang, R. Luo, J. Mao, T. Xiao, Y. Jiang, Acquisition of localization confidence for accurate object detection. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 784-799) (2018). https://doi.org/10.1007/978-3-030-01264-9_48.

[13] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Trans. on Sys. Man, and Cybernetics—Part C: Applications and Reviews, Vol. 34, No. 3, August 2004.

[14] A.J. Lipton, H. Fujiyoshi, R.S Patil. Moving target classification and tracking from real-time video. Proceedings of the IEEE Workshop on Application of Computer Vision, 1998.

[15] T. Weyand, I. Kostrikov, and J. Philbin. PlaNet - Photo Geolocation with Convolutional Neural Networks. In European Conference on Computer Vision (ECCV), 2016.

[16] https://medium.com/@hirotoschwert/digging-into-detectron-2-47b2e794fabd

https://analyticsindiamag.com/detectron2/

[17]https://www.researchgate.net/publication/353531576_Automatic_Object_Tracking_and_Segmentation_Using_Unsupervised_SiamMask

[18] https://link.springer.com/article/10.1007/s11042-022-13644-y

[19] https://arxiv.org/abs/1506.01497