Quest Journals
Journal of Software Engineering and Simulation
Volume 7 ~ Issue 10 (2021) pp: 16-25
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

An Object Detection Scheme in Equirectangular Panaromic Images Using YOLO V3

Dhulipalla Koteswarchand, Moolea Pradeep Kumar, Korapati Naveen Kumar, Md.Sirajuddin
Department Of Computer Science and Engineering
Kallam Harinadha Reddy Institute Of Technology,
Guntur,Andhra Pradesh,India

Abstract - *The main objective of this project is to detect objects in equirectangular paranaromic image using the yolo version 3.the equrectangular panaromic image is formed from 360 degree image i.e A 360 degree images is converted into a 180 degree image or panaromic image. in order to detect an object in equirectangular image first convert 360 degree into equrectangular panaromic and the images are divided into streographic images.thus stereographic images are considered as input for the YOLO V3 algorithm.After coverting into stereographic images the yolov3 model is applied on the image.The yolo v3 includes Conventional neural network for making use of neural network for image processing.this is done by using Darknet-53.ImageAi is used for locating objects in the images.Yolov3 consumes low gpu computing power capable of detecting small objects.Applying commonly available Machine Learning-driven object detection frameworks to stereographic images for object detections.Yolov3 has the best accuracy than the previous.*
Index Terms – *equirectangular panaromic images,object detection,stereographic images,YOLO V3.*

Received 04 October, 2021; Revised: 16 October, 2021; Accepted 18 October, 2021 © The author(s) 2021. Published with open access at www.questjournals.org

I. INTRODUCTION

Panaromic images were first introduced in the year 1845,invented by camera operator to shoot a flim in panaromic mode to cover more objects or actors.The panaromic images are Captured using a Specialized rotating lens in the early 1840 's.Although many improvements Are made to capture a panaromic image using less compoents as possible,due to this developments ,today the panaromic images can be captures using smart phone,a panaromic image can contain more objects than a normal image.Panorama can contain up to the full 360° field of view in all directions around an observer allowing full or partial 3D scenes to be projected onto 2D surfaces. When we capture panoramic pictures, we tend to rotate the camera while taking the picture as we stay in the same position. The camera can be thought of as moving along the curve of a sphere, projecting the spherical image on a resulting plane picture.There are many ways to project spherical images onto a plane. For the case of panoramic images,use of equirectangular projection,which is quite simply the full view around the observer i.e. 360° round and 180° vertically. objects and people in equirectangular panoramas can appear to be distorted to the human eye.

Object detection is an important and challenging field in computer vision, one which has been the subject of extensive research . The goal of object detection is to detect all objects and class the objects. It has been widely used in autonomous driving , pedestrian detection , medical imaging , industrial detection , robot vision , intelligent video surveillance , remote sensing images etc. In recent years, deep learning techniques have been applied in object detection . Deep learning uses low-level features to form more abstractive high-level features, and to hierarchically represent the data in order to improve object detection . Compared with traditional detection algorithms, the deep learning based object detection method based has better performance in terms of robustness, accuracy and speed for multi-classification task. Accuracy and speed for multi-classification tasks. Object detection methods based on deep learning mainly include region proposal-based methods, and those based on a unified pipeline framework. The former type of method firstly generates a series of region proposals from an input image, and then uses a convolution neural network to extract features from the generated regions and construct a classifier for object classes. The region-based convolution neural network (R-CNN) method is the earlier method used to introduce convolution neural networks into the field of object detection. It uses the selection search method to generate region proposals from the input images, and uses a convolution neural network to extract features from the generated region proposals. The extracted features are used to train the support vector machine. Based on the R-CNN method, Fast R-CNN and Faster R-CNN were also proposed to reduce training time and improve the mean average precision. Although region proposal-based methods have higher detection accuracy, the structure of the method is complex, and object detection is time consuming.The latter type of method (based on a unified pipeline framework) directly predicts location information and class probabilities of objects with a single-feed forward convolution neural network from the whole image, and does not require the generation of region proposals and post classification. Therefore,the structure of the unified pipeline framework approach is simple and can detect objects quickly,however, it is less accurate than the region proposal-based approach. The two kinds of methods have different advantages and are suitable for different applications.

A.Motivation

As discussed with in the previous section,Object Detection is a growing technology under the Computer vision based technique.Object dection was being used in many sectors

II. FUNDAMENTALS

360-degree (360◦) video and image content has recently gained momentum due to wide availability of consumer-level video capture and display devices - “Virtual Reality (VR) gear”. Equirectangular panorama has quickly become the main format to store and transmit VR video.ERA images create new challenges for computer vision and image processing as i) we lack annotated 360 datasets for many problems, ii) imagery are often of high-resolution to cover the viewing sphere with reasonable resolution and iii)equirectangular projection creates severe geometric distortions for objects away from the central horizontal line.

In computer vision community, there are several recent works on processing 360 video and images, for example,“compression” of wide angle VR video to conventional narrow angle video. In this project training two state-of-the-art detectors, Faster R-CNN and YOLO (version 3) , with conventional examples available in the existing datasets (ImageNet and COCO in our case) and test them with 360-degree data. In our experiments, the YOLO V3 detector performs better than Faster RCNN, (. To adapt the YOLO V3 detector for less computation power, we propose a multi-projection variant of the original YOLO detector. Our m-p YOLO V3 employs stereographic projection and post-processing with soft non maximum suppression (soft-NMS) and outperforms both Faster R-CNN and YOLO V3.

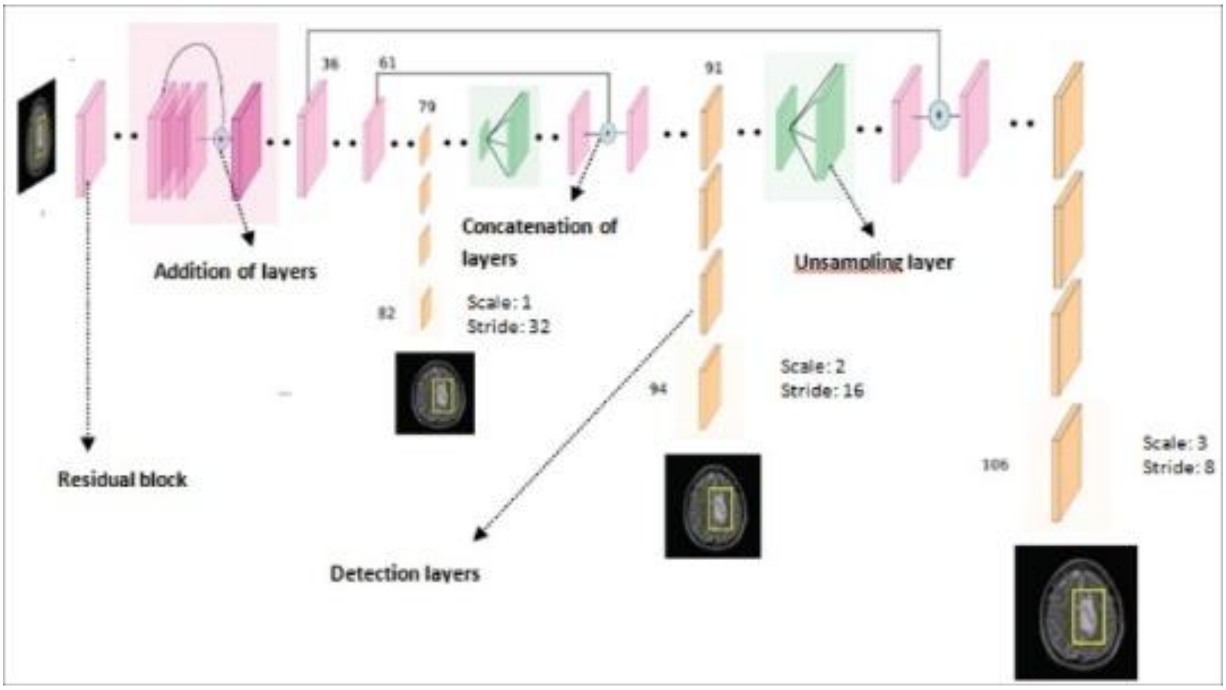

Fig.1. Architecture of YOLO V3 (EUROPE PMC, n.d.)

***A*.DETECTTIVE MODEL**

Detective modeling is used to analyze the given data and detect the outcome. Detctive modeling is used to detect the objects which are already present in the image.In this process,we are going to create,test and validate the model.There are different methods in detecive modelling,they are learning,artificial intelligence and statistics.Once we create the model, we can use many times ,to determine the objects in the images or videos. So,detective model is reusable.Historical data is used to train an algorithm.The detective modelling is an Iterative process .

***B*.PANORAMIC IMAGE TO STEREOGRAPH*C* IMAGES**

Technically, a panorama has an aspect ratio of 2:1 or higher, which means at least twice as wide as its height. The angular extent of a panorama exceeds the typical human binocular field of view of 120° up to a full 360°. From virtual reality to architecture, road infrastructure and urban planning, satellite images of the Earth and even Marscapes make use of this wide image format. As panoramic photography improves in resolution and becomes more and more accessible with consumer devices, the time is right to adapt Machine Learning-based methods to such images.

A bit of geometry will be used to divide a panaromic image into stereographic images inorder To detect objects in the panoramic images. we will move ahead with converting equirectangular panoramas to the stereographic projection.Mathematical formulae are employed in this area inorder to divide a panaromic images into sterograhic images .As the stereographic images will be efficient in detecting.

***C*.YOLO V3 ALGORITHM**

The YOLOv3 method considers object detection as a regression problem. It directly predicts class probabilities and bounding box offsets from full images with a single feed forward convolution neuralElectronics 2020, 9, 537 3 of 11 network. It completely eliminates region proposal generation and feature resampling, and encapsulates all stages in a single network in order to form a true end-to-end detection system.

The YOLOv3 method divides the input image into S × S small grid cells. If the center of an objectfalls into a grid cell, the grid cell is responsible for detecting the object. Each grid cell predicts the position information of B bounding boxes and computes the objectness scores.

3.PROPOSED APPROACH

The main objective of the project is to detetc objects and classify the objects using YOLO V3,the data set which will be used for classifying objects is coco dataset which is mainly used panaromic images.The dataset can be prepared by the use itself or it will be available in coco official site.

At the begining the panaromic images is divided into stereographic images which willbe sent to YOLO V3 algorithm as input,after receriving the images the YOLO V3 starts detecting the images by first applying grid boxes on the images,and then it applies bounding boxes in order to locate and detect object,YOLO V3 gathers confidence of object,using this confidence it classifies the object using machine leaning model.

***A*.Methods**

The proposed method has been implemented YOLO V3 algorithm in order objects in the equirectangular panoramic images.The YOLO V3 algorithm has a total of 106 Convolutional Neural networks and the detection will be carried in three phases at three different neural network layer,at each phase three bounding boxes are applied on the image,here centroiding bounding boxes are used ,i.e.if the centroid of all three boundingboxes are same at each phase then the object is detect ,a machine learning model is used to detect and classify the object.At the beginning a dataset is provided by the user which is useful in labelling the objects,which is available in Coco official website or Kaggle.YOLO V3 is capable of processing 155 frames per second and good for real time processing.Low computational power and less ussage of power.Due to its centroiding bounding box,this model has high accuracy when comapred to previous models.The performance of YOLO V3 is best suited for Panoramic images while comapred to previous versions of YOLO algorithms.

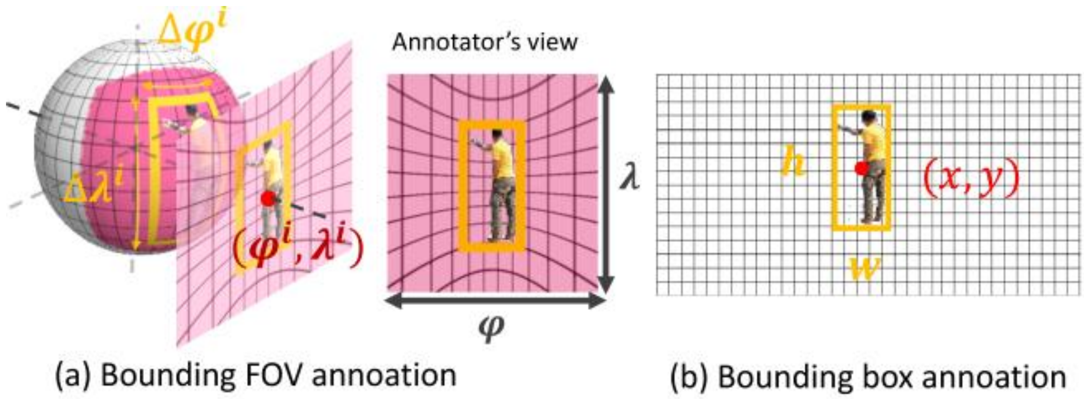

Fig 2. Bounding Box Projections

B.Results

1.Panoramic image to Stereographic images

The input panoramic image will be divided into stereographic images by applying image preprocessing techniques which will used as input for YOLO V3 algorithm to detect objects.

Fig .3**.** An input panaromic image

Fig 2.1 Stereographic image 1.

Fig 2.2 Stereographic image 2

Fig 2.3 Stereographic image 3

Fig 2.4 Stereographic image 4

Here after conversion into stereographic images there are sent to YOLO V3 algorithm
For detecting objects,the output looks like.

2.**Detection of Objects in Stereographic images**

Fig 2.5 Output of YOLO V3 algorithm

Once the stereographic images are fed into YOLO V3 it begins with adding grid boxes on the image,following neural networks,bounding boxes are applied on the images at three phases of total 106 neural networks.

***C*.Performance Of YOLO V3:**

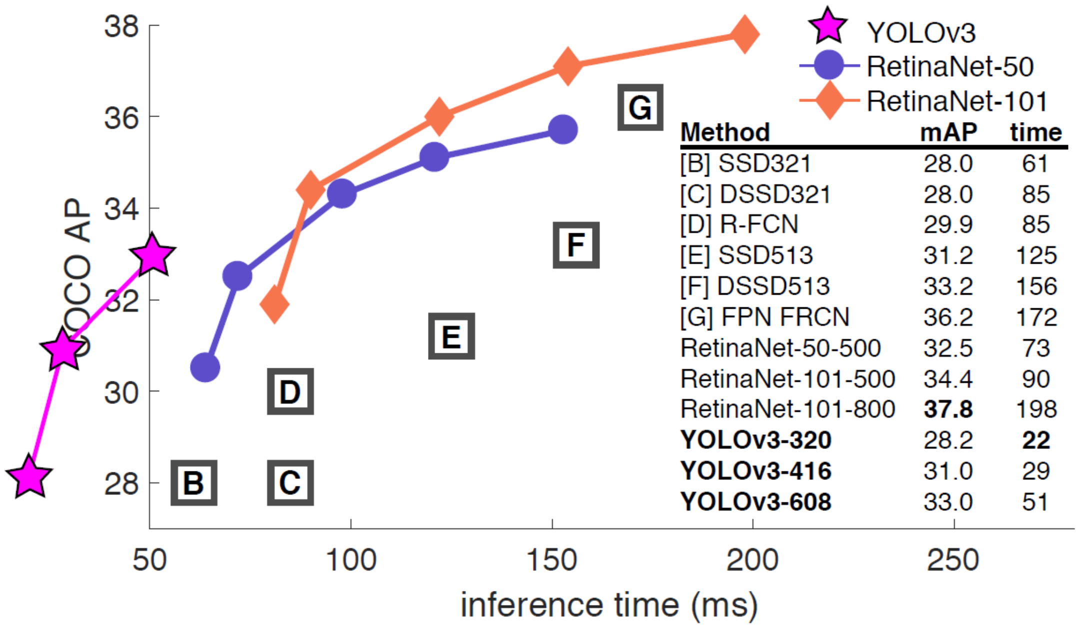

fig.3.We adapt this figure from the Focal Loss paper . YOLOv3 runs significantly faster than other detection methods with comparable performance.

III. CONCLUSION

In this paper, we have applied and proposed to use YOLO algorithm for object detection because ofits advantages. This algorithm can be implemented in various fields to solve some real-life problems like security, monitoring traffic lanes or even assisting visually impaired peoplewith help of audio feedback. In this, we have created a model to detect only three objects which can be scaled further to detect multiple number of objects.

REFRENCES

[1]. Y.-C. Su and K. Grauman, “Pano2vid: Automatic cinematography for watching 360◦ videos,” in ACCV, 2016. 1
[2]. “Making 360 video watchable in 2d: Learning videography for click free viewing,” in CVPR, 2017. 1
[3]. V. Fakour-Sevom, E. Guldogan, and J.-K. Kam¨ ar¨ ainen, “360 panorama ¨ super-resolution using deep convolutional networks,” in Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2018. 1
[4]. U. Kart, J.-K. Kam¨ ar¨ ainen, L. Fan, and M. Gabbouj, “Evaluation of ¨ visual object trackers on equirectangular panorama,” in Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2018. 1
[5]. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Neural Information Processing Systems (NIPS), 2015. 1, 2, 4, 5
[6]. J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in CVPR, 2017. 1, 2, 3, 4, 5
[7]. J. Snyder, Map Projections - A Working Manual. U.S. Department of the Interior, 1987. 1
[8]. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. Zitnick, “Microsoft COCO: Common objects in context,” in ECCV, 2014. 1, 2, 4
[9]. H.-N. Hu, Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun, “Deep 360 pilot: Learning a deep agent for piloting through 360 sports video,” in CVPR, vol. 1, no. 2, 2017, p. 3. 2
[10]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. FeiFei, “ImageNet large scale visual recognition challenge,” Int J Comput Vis, vol. 115, pp. 211–252, 2015. 2