IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

Detection and Content Retrieval of Object in an Image using YOLO

Vinoth Kumar B[1], Abirami S[2], Udhaya R B[3], Bharathi Lakshmi RJ[4], Lohitha R[5]

Associate Professor, Department of Information Technology, PSG College of Technology, Coimbatore, Tamilnadu[1]

UG Scholars, Department of Information Technology, PSG College of Technology, Coimbatore, Tamilnadu, India[2,34,5]

Abstract: It is easy for human beings to identify the object that is in an image. Even if the task is complex, human beings require only a minimal effort. Since computer vision is actually replicating human visual system, the same thing can be achieved in computers when they are trained with large amount of data, faster GPUs and many advanced algorithms. In general terms, Object detection can be defined as a technology that detects instances of object in images and videos by mimicking the human visual system functionalities. The motivation of the paper is making the search process easier for the user i.e., if the object is very new for the user and he has no idea about it, he can upload a picture of that object and the algorithm will detect the object and gives a description about it. The objective of the paper is to detect the object in an image, once the object is detected, the label i.e., the name of the detected object is searched in Wikipedia and few lines of description about that object is retrieved and printed. Also, the label is searched in google and the URL of the top pages with content related to the label are also displayed. The detection of object in an image is done using YOLO (You Only Look Once) algorithm with pre-trained weights. Previous methods for object detection, like R-CNN and its variations, used a pipeline to perform this task in multiple steps. This can take some time for execution, complex optimization may be involved because individual training of components is required. YOLO, does it all fastly with a single neural network. Hence, YOLO is preferred.

Keywords: Object Detection, Region Proposals, Optimization, Yolo, Google Search, Description, Wikipedia, Text to Speech, Artificial Intelligence (AI)

I. INTRODUCTION

Artificial Intelligence (AI) plays an important role in image processing to emulate human intelligence. AI has been used in many image processing fields such as Image Compression[1,2,3,4,5,6,7], Image Segmentation[8,9,10,11], Image Enhancement [12] and Object Recognition [13].It is an easy task for a human being to identify the object that is in an image because of the faster and accurate neural and visual system. Even if a complex task is given, human beings can do it with minimal effort. Since computer vision is actually mimicking human visual system, the same thing can be achieved in computers by training them with large amount of data, faster GPUs and many advanced algorithms. In general terms, Object detection can be defined as a technology that detects instances of object in images and videos by mimicking the human visual system functionalities. The special feature that every object has on its own is used to classify the objects[19]. Example, when searching for circular objects, the objects at a specific distance from the center are searched. Likewise, when searching for square shaped objects, objects are checked for perpendicularity at corners and equality of side lengths. Similarly, for face detection applications standard features like eyes, lips, nose are considered and some other features like skin tone and distance between the eyes are also considered. Due to the circumstances there are some challenges faced during the detection of objects like:

- Lightning: the lightning conditions, weather conditions may vary during the entire course of the day.
- Positioning: the object in the image can be positioned in various aspects.
- Rotation: the object can be in various aspects in the image.
- Occlusion: some part of the object in the image may not be clearly visible.
- Scale: the size of the object may vary.

These are some challenges that should be taken into account while developing an object detection system.

II. ALGORITHMS FOR OBJECT DETECTION

There are several machine learning[23] and deep learning algorithms for object detection. When machine learning approaches are used for detection, it is mandatory to define the features first. Deep learning approaches does not demand to specify the features instead they perform end-to-end detection. Machine learning methods use Support Vector Machine (SVM) and deep learning methods use CNN. R-CNN, fast R-CNN, faster R-CNN are some common algorithms for object detection.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

A. R-CNN

R-CNN uses selective search. By this, bounding boxes are generated. Then, for each bounding box, image classification is done through CNN. Finally, each bounding box are refined using regression[20]. The problems with R-CNN are:

- It takes a huge amount of time to train the network as it requires classification of 2000 region proposals per image.
- It cannot be implemented real time due to the time constraints.
- Since the selective search algorithm is a fixed algorithm, no learning is happening at that stage.

B. Fast R-CNN

In fast R-CNN, the input image is fed into the CNN. Then, the region of proposals are identified and wrapped into squares[17]. Then, reshaping of regions is done using a RoI pooling layer. Then, a softmax layer is used to predict the class of the proposed region[16]. The problem with this fast R-CNN is:

- Performance degradation during testing.

C. Faster R-CNN

- Both of the above mentioned algorithms uses selective search which is a slow and time-consuming process.
- Faster R-CNN does not use selective search. It uses a separate network and by that it produces region proposals[18].
- All of the previously explained algorithms use region based approach to detect the object in the image without looking at the complete image.

D. You Only Look Once (YOLO)

YOLO is also an object detection algorithm which uses only one convolutional network to predicts the bounding boxes and the class probabilities and thus YOLO differs from other region based algorithms[14].

III. WORKING OF YOLO

YOLO trains and tests on full images and directly optimizes detection performance. YOLO model has several benefits over other traditional methods of object detection like the following.

- First, YOLO is extremely fast. Since frame detection in YOLO is a regression problem there is no need of complex pipeline. We can simply run our neural network on any new image at test time to make predictions.
- Second, YOLO sees the entire image during training and testing unlike other sliding window algorithms which require multiple iterations to process a single image.
- Third, YOLO learns generalizable object representations. When trained on real time images and tested, YOLO outperforms top detection methods like DPM and R-CNN.

YOLO network uses features from the entire image to predict each bounding box. It also predicts all bounding boxes across all classes for an image simultaneously. This means our network reasons globally about the full image and all the objects in the image. The YOLO design enables end-to-end training and real time speeds while maintaining high average precision[15].

Following are the steps how YOLO works,

- First it divides the input image into an S × S grid as shown in fig.1.

Fig. 1 Divide the image into S × S grid

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

- If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object.
- Each grid cell predicts B bounding boxes and confidence scores for those boxes as shown in fig.2.
- These confidence scores reflect how confident the model is that the box contains an object. If no object exists in that cell, the confidence scores should be zero.

Fig. 2 Calculate bounding boxes and confidence score for each box.

- Each grid cell also predicts conditional class probabilities.
- These probabilities are conditioned on the grid cell containing an object. We only predict one set of class probabilities per grid cell, regardless of the number of boxes B.
- Finally, we multiply the conditional class probabilities as shown in fig.3 and the individual box confidence predictions which gives us class-specific confidence scores for each box as shown in fig.4.

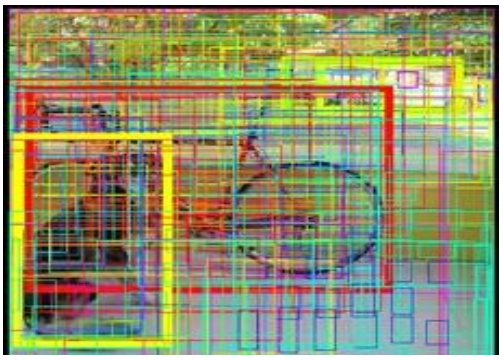

Fig. 3 Multiply probability and confidence scores.

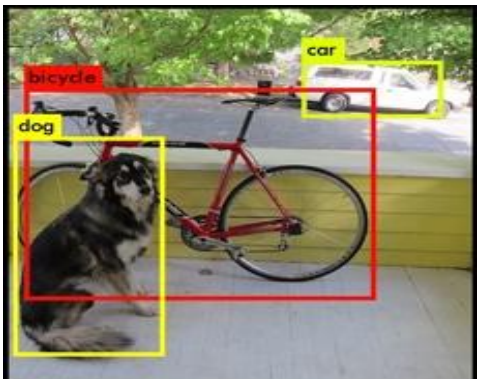

Fig. 4 Final Output

IV. WEB SCRAPING AND TEXT TO SPEECH CONVERSION

Web scraping is a technique that is used to retrieve the content from websites. It consists of two phases namely fetching the web page and later extracting the required content from it. Here two types of web scraping is done one is extracting the content from Wikipedia and other is top google search links for that label. The required modules are installed to system using pip.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

A. Content Retrieval from Wikipedia

After detecting the object from image will use that labelled class to retrieve data from Wikipedia. It is a free encyclopedia in web. So by extracting data from Wikipedia helps the user to get a idea about what the object is and its uses. Wikipedia is a python library that will help to access and extract data from Wikipedia. In that module with a help of a predefined function Summary(), label(object name) and filter(no of lines from Wikipedia) are arguments for this function and returns a string that contains the extracted data.

B. URL Retrieval from Google

By using the label(object name) will extract top google URL's from google with the help of python module Googlesearch. By using pre defined function called Search() will extract the required URL's. In this function we can pass arguments like label(object name) , no of links need to be extracted etc. With these links they can refer more about the object other than Wikipedia content[20].

C. Text to Speech Conversion

This step will convert the label(object name) and Wikipedia content to voice so that everybody can understand better. The module used for text to speech conversion is pyttsx which is platform independent and it can convert in offline too. But pyttsx is supported only in python 2.x versions so pyttsx3 module can used in both python 2.x and 3.x versions. Inorder to use pyttsx3 init() function need to be called to initialize the process and use a predefined method say() with argument text which needs to be converted to voice[19]. Finally use runAndWait() to run the speech.

V. PERFORMANCE ANALYSIS

To analyze the performance of YOLO, it compared with algorithms like R-CNN, fast R-CNN, faster R-CNN on various performance measures like time taken, accuracy and the frames per second. When analysis was done based on time taken by the algorithm to detect the objects as listed in table 1, it is found that R-CNN takes around 40 to 50 seconds, fast R-CNN takes 2 seconds, faster R-CNN takes 0.2 seconds, and YOLO takes just 0.02 seconds. From this analysis it can be inferred that, YOLO performs 10 times quicker that faster R-CNN, 100 times quicker than fast R-CNN and more than 1000 times quicker than R-CNN.

Table I: Performance Evaluation Based on Time Taken

Algorithm	**Time taken** (in sec)
R-CNN	40-50
Fast R-CNN	2
Faster R-CNN	0.2
YOLO	0.02

When analysis was done based on the number of frames per second, YOLO performs far better than all the other algorithms as shown in fig.5, with 48 fps whereas, R-CNN processes 2 fps, fast R-CNN processes 5 fps and faster R-CNN processes 8 fps.

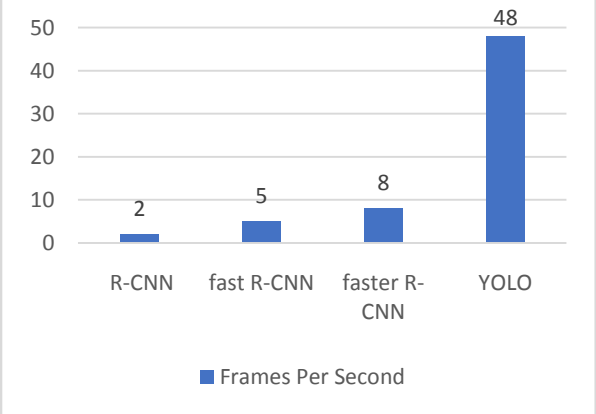

Fig.5 Performance analysis based on frames per second

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

When analysis was done based on the accuracy it is found that YOLO has lesser accuracy than the other three algorithms as shown in fig.6. So, it is not recommended to use YOLO for applications in which accuracy is the major concern.

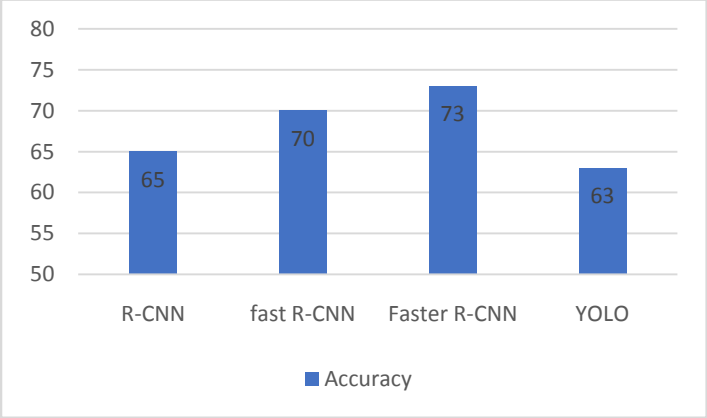

Fig. 6 Performance analysis based on accuracy

The model can be used in tracking objects for example tracking a ball during a football match, tracking movement of a cricket bat, tracking a person in a video, Video surveillance, Smart Class for students, Instructor for blind people to get details about unknown objects. It is also used in Pedestrian detection.

- **Face detection:** An example of object detection in daily life is that when we upload a new picture in Facebook or Instagram it detects our face using this method.
- **People Counting:** Object detection can be also used for people counting, it means that it is used for analyzing store performance or crowd statistics during festivals where the people spend a limited amount of time and other details .This type of analysis is little difficult as people move away from frame.
- **Vehicle detection:** When the object is a vehicle such as a bicycle or car or bus, object detection with tracking can prove effective in estimating the speed of the object. The type of ship entering a port can be determined by object detection based on the shape, size etc. This method of detecting ships has been developed in certain European Countries.
- **Manufacturing Industry:** Object detection is also used in industrial processes to identify products. If we want our machine to detect products which are only circular we can use Hough circle detection transform can be used for detection.
- **Online images:** Apart from these object detection can be used for classifying images found online. Obscene images are usually filtered out using object detection.
- **Security:** In the future we might be able to use object detection to identify anomalies in a scene such as bombs or explosives (by making use of a quadcopter).
- **Medical Diagnose:** Use of object detection and recognition in medical diagnose to detect the X-Ray report, brain tumors.

VI. EXPERIMENTAL RESULTS

Since YOLO is a pretrained model it should have pre-trained YOLO V3 weights file, CFG file,text document containing the object classes in current program directory. This Script needs 4 Arguments.

- Input image – airplane.jpg
- YOLO config file - yolov3.cfg (contains details about layers or hidden layers used in Neural network)
- YOLO pretrained weights file – yolov3.weights (first few layers in Neural network have already learned some general factors applicable to all classes.
- Text file consists of object classes – yolov3.txt (contains object classes names) This model is trained using COCO dataset so it has capability of detecting 80 objects.

Execute the script by this command,

Python yolo1.py--image airplane.jpg--config yolov3.cfg--weights yolov3.weights--classes yolov3.txt

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

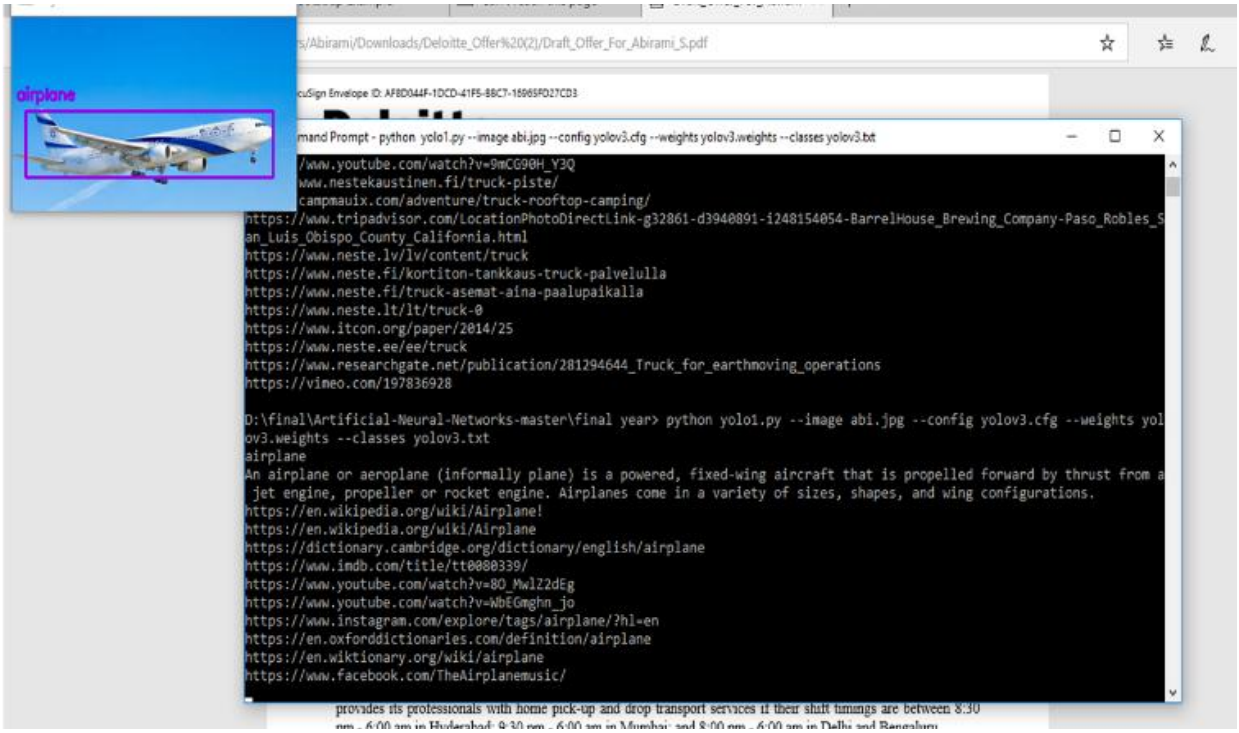

Fig.7 Airplane

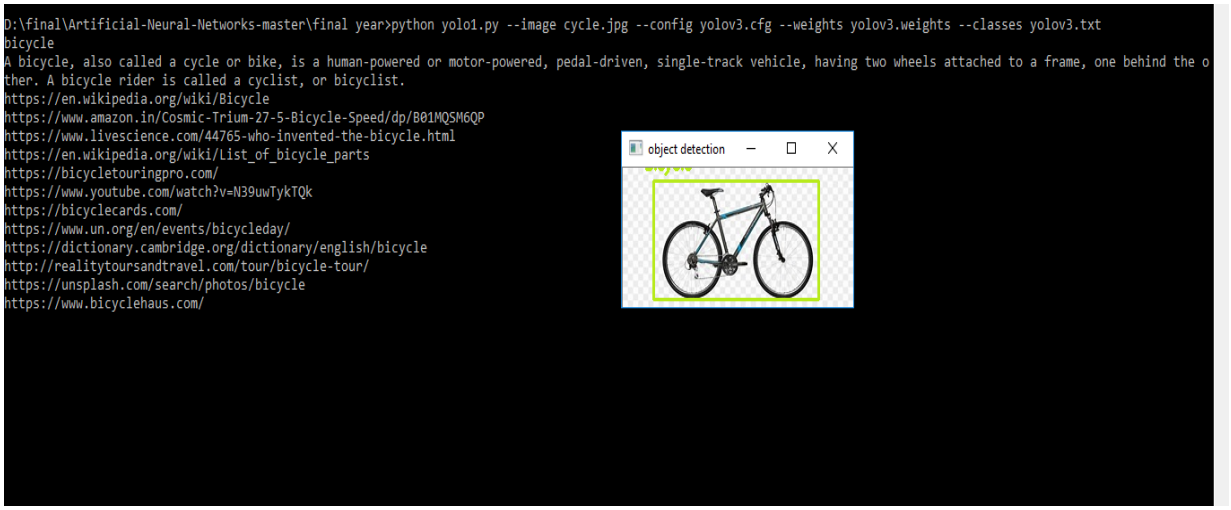

Fig.8 Bicycle

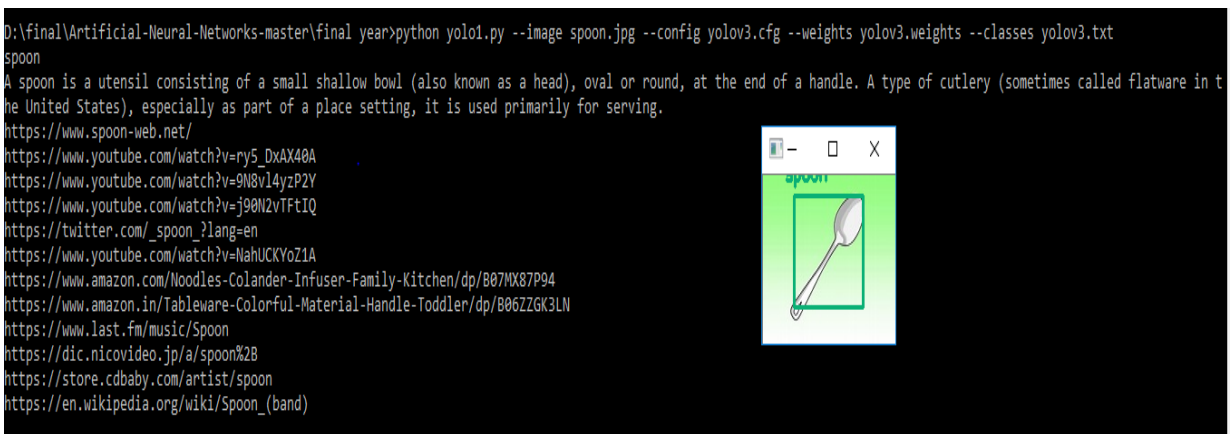

Fig.9 Spoon

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 6, Issue 6, June 2019

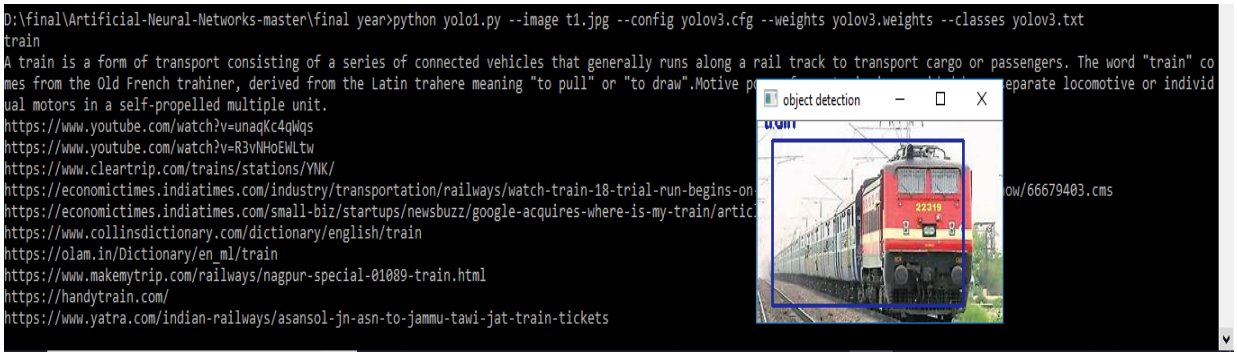

Fig.10 Train

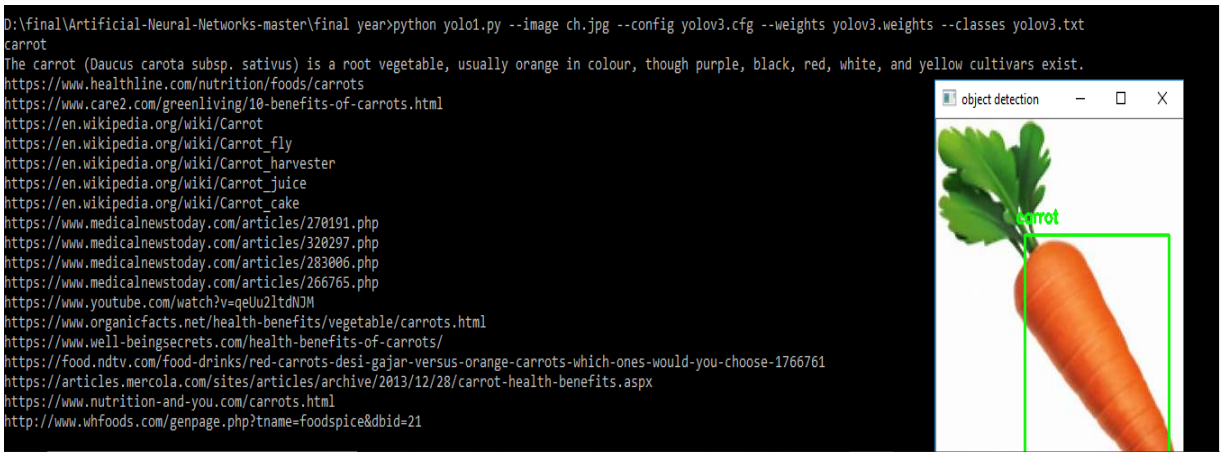

Fig.11 Carrot

Label (object class) will be printed and it will be converted to voice and then the content retrieval from Wikipedia gets executed as shown in fig.7,8,9,10 and similarly it will be converted to voice. Now the top Google links will be shown finally the image with the detected class label is seen with bounded box around the object. After closing this window the image with labelled object will be saved in the program directory.

Although the results are encouraging, the model has few limitations as follows:

- YOLO has strong spatial constraints and hence it cannot detect small objects which appear in groups.
- A small error in small grids can cause a greater impact in the result.
- The model struggles when object is in new aspect ratio or configuration.

V. CONCLUSION

An efficient model is developed in this research work that generates the audio of the descriptions for objects present in the given image. Identification of objects is achieved through YOLO and the descriptions for the detected objects are generated using Wikipedia package and the related URLs are retrieved using google search packages available in python library. The fetched descriptions are read out using the pyttsx3 package. Future avenue is to enhance this to an image based search engine.

REFERENCES

[1]. B.Vinoth Kumar, G.R.Karpagam, "A Smart algorithm for Quantization Table Optimization: a case study in JPEG Compression" Smart Techniques for a Smarter Planet: Towards Smarter Algrthm, "Studies in Fuzziness and Soft Computing" book series, Springer, 257-280, 2019.

[2]. B.Vinoth Kumar & G.R.Karpagam, 2018 "Reduction of Computation time in Differential Evolution based Quantization table optimization for the JPEG baseline algorithm" International Journal of Computational Systems Engineering, Inderscience Publishers Vol.4 No.1 2018, pp.58-65.

[3]. B.Vinoth Kumar, G.R.Karpagam," A Problem Approximation Surrogate Model (PASM) for fitness approx in optimizing the quantization table for the JPEG baseline algorithm" Turkish Journal of Electrical Engineering and Computer Sciences, 24 (6) (2016) 4623-4636.

[4]. B.Vinoth Kumar and G.R.Karpagam, (2017) "Single Versus Multiple Trial Vectors in Classical Differential Evolution for Optimizing the Quantization Table in JPEG Baseline Algorithm", ICTACT Journal on Soft Computing Vol. 7 No.4 2017, pp.1510-1516.

[5]. S.P. Naresh, B.Vinoth Kumar and G.R.Karpagam, (2015) "A Literature Review on Quantization Table Design for the JPEG Baseline Algorithm", International Journal of Engineering and Computer Science, Vol. 4, No. 10, 2015, pp. 14686-14691.

[6]. S.Viswajaa, B.Vinoth Kumar and G.R.Karpagam, (2015) "A survey on Nature inspired Meta-Heuristics Algorithms in Optimizing the Quantization Table for the JPEG Baseline Algorithm. International Advanced Research Journal in Science, Engineering and Technology, Vol. 2, No. 4, 2015, pp. 114-123.
[7]. B.Vinoth Kumar and G.R.Karpagam, (2011) "An Empirical Analysis of Requantization Errors for Recompressed JPEG Images", International Journal of Engineering Science and Technology, Vol. 3 No.12 2011, pp. 8519-8527.
[8]. B.Vinoth Kumar, G.R.Karpagam, and Yanjun Zhao "Evolutionary Algorithm with Memetic Search Capability for Optic Disc Localization in Retinal Fundus Images" Intelligent data analysis for biomedical applications: Challenges and Solutions, "Intelligent Data-Centric Systems" book series, Elsevier, 191-207, 2019.
[9]. S. Bharkad, Automatic segmentation of optic disk in retinal images, Biomedical Signal Processing and Control 31 (2017) 483–498
[10]. B.Vinoth Kumar, Janani K and N MythiliPriya, A Survey on Automatic Detection of Hard Exudates in Diabetic Retinopathy, IEEE International Conference on Inventive Systems and Control, January 19-20, 2017
[11]. B.Vinoth Kumar, G.R.Karpagam and N.Vijaya Rekha," Performance Analysis of Deterministic Centroid Initialization Method for Partitional Algorithms in Image Block Clustering", Indian Journal of Science and Technology, Vol 8(S7), 63–73, April 2015.
[12]. Sushil Kumar & Millie Pant & Amiya Kumar Ray, 2018. " DE-IE: differential evolution for color image enhancement," International Journal of System Assurance Engineering and Management, Springer; vol. 9(3), pages 577-588.
[13]. Zhong-Qiu Zhao, PengZheng, Shou-taoXu, Xindong Wu , Object Detection with Deep Learning: A Review, IEEE Transactions on Neural Networks and Learning Systems, 1-21, 2019.
[14]. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali 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.
[15]. Impiombato, D., S. Giarrusso, T. Mineo, O. Catalano, C. Gargano, G. La Rosa, F. Russo et al. "You Only Look Once: Unified Real-Time Object Detection." Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 794 (2015): 185-192.
[16]. Wang, Xiaolong, Abhinav Shrivastava, and Abhinav Gupta. "A-fast-rcnn: Hard positive generation via adversary for object detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2606-2615. 2017.
[17]. Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.
[18]. Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems, pp. 91-99. 2015.
[19]. Vondrick, Carl, etal. "Hoggles: Visualizing object detection features." Proceeding of the IEEE International Conference on Comp Vision. 2013.
[20]. Beena, M. V., MN Agnisarman Namboodiri, and P. G. Dean. "Automatic sign language finger spelling using convolution neural network: analysis." International Journal of Pure and Applied Mathematics 117.20 (2017): 9-15.
[21]. Manaswi, Navin Kumar. "Speech to Text and Vice Versa." deep learning with applications using python. Apress, Berkeley, CA, 2018. 127-144.
[22]. Bharanipriya, V., and V. Kamakshi Prasad. "Web content mining tools: a comparative study." International Journal of Information Technology and Knowledge Management 4.1 (2011): 211-215.
[23]. I.Devi, G.R.Karpagam and B.Vinoth Kumar, (2017) "A Survey of Machine learning Techniques" International Journal of Computational Systems Engineering, Inderscience Publishers, Vol. 3 No.4 2017, pp.203-212.