I. J. Engineering and Manufacturing, **2022, 2, 34-40**
Published Online April 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2022.02.04

Guiding Aid for Visually Impaired

Pragati Chandankhede
SPSU, Dept of Computer Engineering, 313001, India
Email: pragati.chandankhede@spsu.ac.in

Arun Kumar
SPSU, Dept of Computer Engineering, 313001, India
Email: Arun.kumar@spsu.ac.in

Received: 28 December 2021; Revised: 20 January 2022; Accepted: 04 February 2022; Published: 08 April 2022

Abstract: Visual impairment is where the person either can't see or his vision has weakened to large extent. There is no alternative technique for visually impairment, but to some extent it can be trim down with devices, smart sticks and sensors. Although many techniques are there for helping out through electronic travelling aid, cost effective and minimum hardware solution was the expectation by impaired. The device which can identify and classify the object ahead of impaired person is needed so that person can be prevented from the accident. In this paper, a unified model of YOLO (You Only Look Once) is used for detection of object ahead of camera. The proposed model is based on phenomena of detecting small object and good detection speed of yolov3 makes system more robust. Once detected, labeled objects name is converted from text to speech, so that blind person can be alerted from colliding with obstacles. This paper is one step in the direction to help them by exactly classifying, detecting and localizing target object along with providing voice based guideline. The proposed model has proved accuracy in many real time scenes.

Index Terms: Computer vision, object detection, electronic aid, feature extraction, object labeling

1. Introduction

There are about 285 million people who are visually impaired, out of which 246 million people have low vision and 39 million are completely blind, the source to data is World Health organization. When it comes to Visual impairment, many technologies have come forward for helping; it can be broadly classified as adaptive system and assistive technology. The purpose of Adaptive system were to react to the environment in which impaired person is in and guide them for activity. While assistive technology enable people assist current environment. 3 D sound maps, Braille display are few benchmark of assistive technology [5].

The major contribution is enlightened herein. Author C. Wong, D. Wee, I. Murray and T. Dias, in year 2001 has proposed a system that centre's around the use of a micro controller to calculate distance measurements of ultrasonic signals sent and received by two sets of transducers fixed onto the white cane[12]. The difficulty was faced by the user at the point, when travelling in crowded places the cane should be kept close to the body to avoid tripping other people, other problem faced on It cannot warn the user of head level obstacles, drop offs, and obstructions over a meter away. Than in 2003, Jeremy Hill and Black Jones, the device popularly known as MiniGuide US was developed that uses ultrasound to detect objects, and gives tactual or auditory feedback by vibrating or chirping more rapidly as you approach an object. Five default ranges: 8 meter, 4 meter, 2 meter, 1 meter, and 1/2 meter. It still reside as a good and helpful aid , but can't be considered as primary aid for navigating user and cannot rely completely on it[13].

Maymounah Alshajajeer, Maryam T. Almousa in year 2018 proposes System that consist of Ultrasonic Sensor, Raindrops Sensor Module, LDR (Light Dependent Resistor) or Photocell, Buzzer, these component combine help visually impaired person to locate object. The difficulty faced was the returning signal from the ground and the immediate lack of returning signal from the drop-off is very small[14].

Various travelling aid have been developed that are capable to give information on obstacles in the scene and hence help to avoid them. The problem that resides with the system was majority of them are dependent on ultrasound or sensor based technology which suffer from the problem of elevation problem. If the blind person is o steep way the sharp changes in elevation of sensor or ultrasound can't be detected exactly and user will be misguided. Another reason is all the system is go-no-go system. In this paper We develop a framework which exploit by computer vision[10].

The area of Computer Vision have made user possible to see the content of digital images. Identification of digital images completely depends on feature extraction.[2] The challenge of feature extraction is to understand the useful data and data of interest among data available to system at given instance. Along that it's required to have labeled data for

being identify feature. Many algorithms have come forward to overcome this challenge. Convolution Neural network(CNN) is the foundation for feature extraction, with the help of convolution and pooling layer and siding window approach it made performance possible to higher object detection[1]. Image classification was possible with Deep convolution network (DCNN) because of fully connected layers that identify exact output class. In order to identify different features at each location, feature maps are generated.

The approach for real time object detection is not a dream today, the technique of R CNN, Yolo, Coco model have made a remarkable revolution for real time detection[6]. Aim of researcher is to prepare learning algorithm and make the algorithm understand scene present in front of them ie, allowing machine to model our world[7]. To prepare algorithm seems being simple but second part to let them understand scene and react is quite difficult. It's because raw input to such learning system is always high dimensional entity. It may be solid object with 3Dimensional view. For such a case, computer vision and object recognition work together. Image recognition is best handled by deep learning technique of AI, which can better serve visually impaired person.

Before the emergence of innovative approach in object detection, the classifiers are used to perform detection at different location and at different scales[9]. This approach is based on selection of interested part in image. This is consider as an detection with regression approach. This approach is expensive while processing. Hence the approach was needed that would directly evaluate the whole image at once. And the YOLO techniques help us in doing so. YOLO: was considered to be an impactful algorithm that identify object in all the sense, is released at 2016.

2. Choosing Yolo to Mold the System

2.1 Choosing Yolo

Various object detection systems come into existence, but the problem with many of them was, they depend on classifiers to be applied at multiple scale and location. The open source Neural Network, Darknet has presented many model, You only look once is one of them[8]. Yolo resolves the problem by applying single network to full image. The network divides the image into regions and predicts the possibilities of object in an image with help of bounding box. The interesting part of Yolo v3 is its small feature map that packs lots of information within it. It can also be said as loss function, which has the capability of predicting same object of different size. The feature of IoU, Intersection over union value that notify that whether the object is present in said bounding box or not, it ranges 0 if not present and 1 if its perfectly able to predict. The purpose of loss function is to find most excellent IoU.

2.2 Choosing Yolo V3

The aim of object detection algorithm is classification and localization. To determine whether the algorithm does it depends upon mean average precision said as maP. Its value that holds for correct classification or correct percentage where the system correctly predicted the object. The detection algorithm performance decides on maP. Another factor that can be used to judge the algorithm is how much frame per second the algorithm detects. YOLO initial performance was at 45 frames per seconds for real time detection. Yolo V2 performs 67 frames per seconds' detection. YOLO Version 3 proved to be more successful with tiny fine prediction. While YOLOv3 predicts three bounding boxes per cell (compared to 5 in YOLOv2) but it makes three predictions at different scales, totaling up to 9 anchor boxes. Performance of Yolo can be judged, with the precision of detected bounding box. The paper focuses on verion 3 Yolo which best works and yields accuracy on real time model. The important parameter of Yolo is detection accuracy, which it does by dividing the input image into a size of $S \times S$ grid. The grid is then responsible for detection of object. Center in a grid cell enlighten, object is detected. Yolo is being trained on COCO data set which primarily has the ability to classify object out of 80 categories. The boundary that identifies the object is called as bounding box. Its always better to use pre trained model for identifying the bounding box. Its not feasible to write a code for every application. Various third party implementations that are designed using YOLO can be used for object detection. To perform object detection and output a model, the program can be written that can use pre trained weights. Yolo provides the capability of using pertained version which is helpful for many real time projects. One of reason to make use of pre-trained models is to be able to use images that are readily made available and get easy with the framework.

2.3 Working of Yolo V3

It's interesting to know about the detection of Yolo works.

1. First step is to division of given image into fixed size dimension S*S, Usually the Bounding Box are of fixed dimension.
2. The object resides in grid are detected by the algorithm. Each grid cell provides a confidence score; this is the value that indicates object is present in box. As shown in Fig.1.

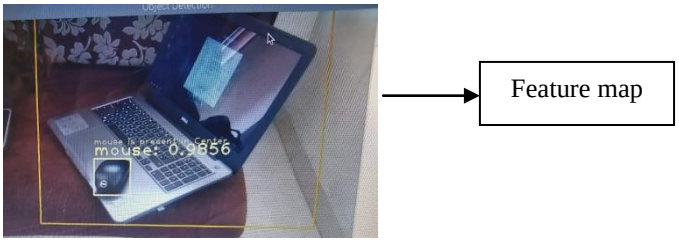


Fig.1.Presence of Object

3. Once object is detected the bounding box is enclosed within it. There is certain feature that decides the object. If input image is of laptop/ computer/ TV then the screen will decide its feature map, while other feature map can be presence of keyboard / screen size etc. that is the feature map is extraction of useful feature within the bounding box.
4. Next important thing to know about bounding box, it has following three attributes with it, width, height, class, bounding box coordinates(x,y) (Tw, Th, bx, by). Yolo v3 predicts boxes at three different scales.
5. The class of detection is trained by coco data set.
6. In YOLO v3 Logistic regression is responsible for calculating object confidence and class predictions.
7. Next step is to identify whether the image consist of two or more detected objects.
8. IoU (Intersection over union) is calculated to understand performance of detection, it indicate the ground truth of object present and detected bounding box.

IoU= Area of overlap / Area of union.

As the value lies between 0 and 1. 1 indicates the perfect presence of object in bounding box. In case of multiple object detection, the IoU value is expected to be greater than 0.5. As shown below, every box here indicate the object.

i) The IoU value of the above box indicate the value of presence of one or more object. But due to false classification it can hold four different values.
ii) IoU >=0.5, it will classify the object as True positive.
iii) IoU<0.5, it will classify the object as False positive.
iv) If the system fails to identify object it classified it as false negative.
v) If no object is beig predicted by model, it classify as True Negative.

9. If object is identified, text to speech converter is used, object ahead is being read by the system for blind person.
10. Location of object is alerted to person by guiding user with center, bottom left, bottom right, upper left, upper right.

3. Data Set

The ImageNet dataset consist of high resolution images of about 15 million images. And it has about 22 thousand categories. With the approach of deep learning, huge data set is required. YOLO has been pretrained on COCO data set[7]. Yolo classify object with ImageNet samples for each COCO data. COCO data set has 80 classes.

4. The Architecture

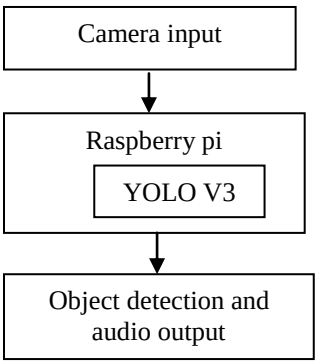


Fig.2. The architecture of the system

The figure 2 illustrate the overall architecture that has been used to build the system.The details have been showcase as below.

1. Raspberry Pi model B

The current era depends on compact size architect devices. This gives a chance to user to create something innovative. The Raspberry pi is computer of shape as small as debit card, the performance has built-in due to quad core ARM processor. It has four USB slots that can be used for multiple purposes.

2. Yolo V3

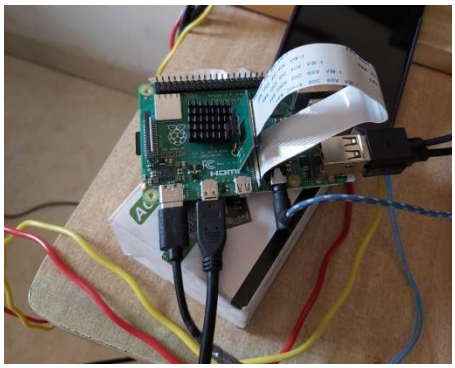
To perform object detection and output a model, the program can be written that can use pre trained weights. Performance of Yolo can be judged, with the precision of detected bounding box.

3. Audio device.

Audio device input of headphone is attached to raspberry pi, once object detected, the enclosed bounding box is created by Yolo v3, it categorize object and given as audio output for user.

Innovation as a device: Embedding the Yolo v3 in Raspberry pi module can make the detection useful for visually impaired as it can be mounted in hand by the user or in shirt pocket. Choosing Yolov3 as strong detector since it get boxes perfectly aligned with the object detected. This allows free navigation for visually impaired.

5. Result and Observation



a

Fig.3.a The system with all attached components

b

Fig.3.b. The system as a Device

In the above figure 3.a the raspberry pi is connected with power supply at 5A/3V DC. Camera of 5 Mp is attached to pi module for capturing images. It's one of the USB is also being connected to monitor, where we can watch the output images with bounding boxes. This monitor is not thoroughly applied to system, its being connected to simply visualize the result in testing phase. There is an attachment of headphone which can give audio output to the user. There is no such Region of Interest in an image that needs to be filtered out for processing. Typically the entire detected object by the system will be processed using COCO data set. The complete system is assembled at a place as shown in figure 3.b. Assuming the fact that the visually impaired maintain the peace while walking, they do not and cannot walk fast. Hence step that will be followed while operating this system is:

1. Switch on the system when need to be used.
2. Mounted camera will start recognizing image from the frame captured by camera.
3. The recognized image will be labeled by coco data set.
4. This labeled image will be read by the system.

5. The reading can be heard by the headphones attached to the system.
6. The information about location of the detected object will be told to user, from where user can make further movement.

Fig. 4. Screenshot of Multi object detection and object location identification

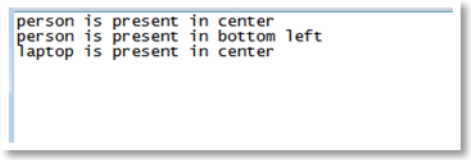


Fig. 5. Screenshot of result that can be visible on desktop.

As shown in the figure 5, With the center of attention of camera module, location of object is alerted to person by guiding user with center, bottom left, bottom right, upper left, upper right.

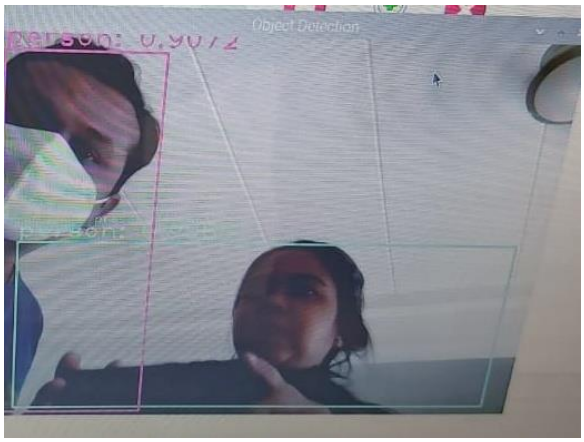


Fig.6. Multiple face detection identified as Person.

Recognizing faces from the captured image of raspberry pi camera is not treated apart from object detection; its part of object detection, yolov3 with the applicability of widerFace dataset can recognize almost 393703 face labels. The figure 6 illustrate that the system is able to recognize multiple face showing accurate objectness/ person score that inform the probability of person being present in captured image.

6. Comparison and Discussion

Darknet is Open source neural network framework that was proved as base for Yolo. The proposed methodology uses yolov3. YOLO v3 make use of a approach that can identifies multilabel and allows detection of specific classes and have numerous for individual bounding boxes. Despite of various approach of yolo v3 is choosen as it tiny verson make model fits best in Pi device and up to mark performance can be achieved. The Model accuracy calculated by comparing the ground truth of actual object presented within the scene with the predicted object detected by the system. Short summary of different classification results are detailed herein. The Yolo V3 and Yolo v4 version have proved efficient as the recognition of was more accurate. Mask RCNN & Mobile Net SSD are technique for object detection. In

term of speed of detection, the result was in support of Yolo which surpass Mask R-CNN with almost 20 times. In term of Size, yolo make use of input size of 416 x416, while the input image size to the Mask R-CNN is 1024x1024. While 512x288 size of images are processed by Mobile net SSD. The short description of the object detection result for common object is mentioned below.

Table 1. Comparison of model

ImageName	Actual Objects Present (groundtruth)	Recognized by Yolov3	Recognized by Yolov4	Recognized by MaskRCNN	Recognized by MobileNetSSD	Yolov3	Yolov4	MaskRCNN	MobileNetSSD
bird.jpg	8	8	7	7	3	100.00%	87.50%	87.50%	37.50%
buses_cars.jpg	6	6	6	5	4	100.00%	100.00%	83.33%	66.67%
cars.jpg	16	15	16	9	7	100.00%	100.00%	56.25%	43.75%
dog-cat.png	2	2	2	2	2	100.00%	100.00%	100.00%	100.00%
dog-cat1.jpg	4	4	4	4	4	100.00%	100.00%	100.00%	100.00%

7. Performance of Our System

Precision and Recall are the point of calculating performance of the system.

As discussed in section TP (true positive), TN (true negative), FP (false positive), FN (false negative) are the factors.

The true positive is the situation when object that model predicted was true to that of ground truth.

The True negative is the situation that no object presents is predicted and it was the ground truth too.

The false positive is situation when model predicts thing which is not ground truth, neither present in scene.

The false negative is the situation of predicted object is not part of scene, but that's not the ground truth.

Real time testing the system gives following result.

Here TP =90 indicates, no of time laptop was predicted correctly.

And FP = 02 indicates, no of time laptop was not predicted as it was not the part of scene.

TN= 00 indicates, no object was presented scene and same is predicted by the system.

FN=02, object is not predicted by the system, ground truth was there was object in scene.

$$\text{Precision} = (\text{True Positive})/(\text{True Positive} + \text{False Positive}) \quad (1)$$
$$= 90/(90+2)$$
$$=97.82$$

$$\text{Recall}=\text{True positive}/ \text{True positive} +\text{False Negative} \quad (2)$$
$$=90/ (90+25)$$
$$=0.78$$

The precision value is on top of recall in the above observation, since there are imbalanced classes in real time detection.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$
$$=(90+02)/ (90+00+02+02)$$
$$=92/ 94$$
$$=97.87\ \%$$

The system guarantees best real time detection in minimum time with 97.87% accurate. The raspberry pi model with 8gb ram and the backbone of Yolov3 has express the computation that has satisfied result as envision by visually impaired. The system necessitates 3.7 seconds to scan orevall scene. There is no simple technique that can improve performance of real time detection, it's obvious that adding classifier will definitely improves performance but at same time it improves time for computation. The system can reduce false detection rate and blind can be alerted via voice commands and navigation can be completed effortless.

8. Conclusion

The object detection phenomena of computer vision have made a remarkable process of understanding digital images. The architecture explains in the paper is one step in direction to assist visually impaired. There is applicability of Yolo V3 capability of predicting class and guiding accurate prediction in a captured image. The cost effective

solution using Raspberry Pi processor has make the system more robust. The system accuracy is 97.87 % with an assumption that blind move ahead without support with slowly pace. With this assumption system proves best processing of real world mages and notifying user about the object ahead with audio output. Through which user can be alerted. I conclude that the system is best solution for obstacle detection so that blind can move independently even to unknown place. Limitation of the system is even though the classification and localization is the priority of this type of application, at same time prediction time should be minimal. Second limitation that lies here is the detected object aspect ratio. Its important to guide impaired person about size of object presented ahead (somehow it depends on camera angel and object nearer to camera will be treated as big size, that may not be reality). In future this program can be used to guide distance from user to object to collide from obstacles.

References

[1] Karen Simonyan_ & Andrew Zisserman, "Very Deep Convolution Networks for Large-Scale Image Recognition" ICLR 2015.

[2] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors" , arxiv, 2012.

[3] Shraga Shoval, Johann Borenstein, and Yoram Koren, "Mobile Robot Obstacle Avoidance in a Computerized Travel Aid for the Blind" IEEE International Conference on Robotics and Automation, 1994.

[4] Shuihua Wang , Xiaodong Yang , Yingli Tian, "Detecting signage and doors for blind navigation and wayfinding" Network Modeling Analysis in Health Informatics and Bioinformatics , 2013.

[5] Filippo L.M. Milotta, Dario Allegra, Filippo Stanco, Giovanni M. Farinella "An Electronic Travel Aid to Assist Blind and Visually Impaired People to Avoid Obstacles" Springer International Publishing Switzerland. 2015.

[6] Ross Girshick, "Fast R-CNN", IEEE International Conference on Computer Vision, 2015.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun,"Deep Residual Learning for Image Recognition" IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016.

[8] Joseph Redmon Ali Farhadi , "YOLOv3: An Incremental Improvement" , arxiv , 2018.

[9] Fares Jalled, Ilia Voronkov "Object Detection Using Image Processing", arxiv, 2016.

[10] Karen Simonyan, Andrew Zisserman, "Very Deep Convolution Network for Large Scale Image Recognition", ICLR 2015.

[11] Piotr Dollár, Pietro Perona, Serge Belongie "The Fastest Pedestrian Detector in the West", British Machine Vision Conference, 2010.

[12] C. Wong, D. Wee, I. Murray and T. Dias ,"A Novel design of Integrated Proximity Sensors for the White Cane" , The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001.

[13] Jeremy Hill and Black Jones, "The Miniguide: A New Electronic Travel Device", Journal of Visual Impairment and Blindness, Volume 97 issue 10, 2003.

[14] Maymounah ALSHAJAJEER, Maryam Taqi Almousa, Qasem Abu Al-Haija, Journal of Applied Computer Science & Mathematics "Enhanced White Cane for Visually Impaired People" Issue 2, (Vol. 12), 2018.

[15] Htwe Pa Pa Win, Phyo Thu Thu Khine, Khin Nwe Ni Tun "Face Recognition System based on Convolution Neural Networks" International Journal of Image, Graphics and Signal Processing, MECS, 2021.

[16] Md. Rezwanul Haque, Md. Milon Islam, Kazi Saeed Alam, Hasib Iqbal, Md. Ebrahim Shaik "A Computer Vision based Lane Detection Approach" International Journal of Image, Graphics and Signal Processing, MECS, 2019.

Authors' Profiles

Pragati Chandankhede is PhD scholar from Sir Padampat Singhania University Rajasthan. She has completed her MTech from G.H.Raisoni College of Engineering in Year 2013. She has been Engineer in stream of Information technology in year 2009 from JDIET, Yavatmal.

Pragati Chandankhede is Assistant Professor at KCCEMSR, Thane. She has 10 years of experience in teaching. Her area of interest includes artificial intelligence, Natural Language Processing, soft computing, software Engineering. She is PhD Pursuing from Sir Padampat Singhania University Rajasthan. Her PhD works mainly focuses on Computer vision and digitization of real world images. Her recent published article was titled "Using Machine Learning for Image Recommendation in News Articles" published in Recent Advances in Artificial Intelligence and Data Engineering pp 215-225 on November2021. Another published article was titled "Gesture Based Media Controlling using Haar Cascade" published in International Conference on Innovative Computing and Communications pp 541-551 on August 2021. Another published article titled "Deep Learning Technique for serving Visually Impaired Person" in IEEE Xplore on 14 May 2020.

Pragati Chandankhede is member of Professional Society of IETE and ISTE.

How to cite this paper: Pragati Chandankhede, Arun Kumar, " Guiding Aid for Visually Impaired", International Journal of Engineering and Manufacturing (IJEM), Vol.12, No.2, pp. 34-40, 2022. DOI: 10.5815/ijem.2022.02.04