Engineering and Technology Journal
Journal homepage: https://etj.uotechnology.edu.iq

Blind Assistive System Based on Real Time Object Recognition using Machine Learning

Mais R. Kadhim *, Bushra K. Oleiwi

Control and Systems Engineering Dept., University of Technology-Iraq, Alsina'a street, 10066 Baghdad, Iraq.
*Corresponding author Email: 60684@student.uotechnology.edu.iq

HIGHLIGHTS

- A system that helps blind people discover the objects around them using one of the deep learning algorithms called Yolo.
- The system consists of two parts: the software, represented by the Yolo algorithm, and the hardware part is the Raspberry Pi.
- The proposed system is characterized by high accuracy and good speed.

ARTICLE INFO

Handling editor: Muhsin J. Jweeg
Keywords:
Objects detection
Open CV
Python
YOLO
Deep neural network

ABSTRACT

Healthy people carry out their daily lives normally, but the visually impaired and the blind face difficulties in practicing their daily activities safely because they are ignorant of the organisms surrounding them. Smart systems come as solutions to help this segment of people in a way that enables them to practice their daily activities safely as possible. Blind assistive system using deep learning based You Only Look Once algorithm (YOLO) and Open CV library for detecting and recognizing objects in images and video streams quickly. This work implemented using python. The results gave a satisfactory performance in detecting and recognizing objects in the environment. The results obtained are the identification of the objects that the Yolo algorithm was trained on, where the persons, chairs, oven, pizza, mugs, bags, seats, etc. were identified.

1. Introduction

Visually impaired people have a difficult time moving around safely and independently, which prevents them from participating in uniform professional and public activities both indoors and out. Similarly, kids have had a lot of practice identifying the principles of the surrounding environment. According to a WHO (World Health Organization) statistical analysis research, roughly 285 million people are blind or have amblyopia around the world, with 246 million having major vision impairment [1]. Blind people have a tough time interpreting their surroundings, which is one of the most serious issues. One of the biggest problems is that blind people face difficulty understanding the environment around them. Blind people depend in their daily lives on other people, guide dogs, or electronic devices. Object detection is one of the primary tasks in the field of computer vision. The ideal solution to address this problem is to train the object detector that works on a specific part of the image and then apply these discoveries in a very fast way; this method achieved high success due to its maximum speed and high accuracy. Recently, several techniques have been proposed to help blind and visually impaired people discover objects around them [2], [3]. For example, some researches have relied on artificial intelligence techniques, others have used ultrasound signals, others have used deep learning, and there are a lot of researches dealing with helping people with visual impairment to discover objects surrounding them and avoiding obstacles. Several featured work has been advanced in the form of blind sailing systems: voice system [4]. With the help of a GPS system, a portable computer, and laser inputs, traditional eyeglasses are connected to a camera by which things in video images are recognized and converted to sound. The Tiflis prototype [5] comprises of two cameras, a sensor attached to black glasses, and a vibration array that informs the user [6], [7].With the help of a GPS gadget, a laptop computer, and RFID (Radio Frequency Identification) technology [8]. XASBliP (Cognitive Aid System for Blind People) [9] is another system that includes a pair of glasses and a helm, as well as a tiny laptop and an FPGA hand. In order to tackle object detection difficulties, several navigation systems rely heavily on machine learning [10].Researchers frequently employ classification techniques that are close to global test model optimization, such as

http://doi.org/10.30684/etj.v40i1.1933
Received 21 November 2020; Accepted 13 February 2021; Available online 25 January 2022
2412-0758/

SVMs (Support Vector Machines) and Adobos (Adaptive Boosting), which have been widely employed in vision applications [11].To detect things such as vehicles and persons, SVM (Support Vector Machine) has been utilized to sort hear wavelet coefficients describing Gabor and edges properties [12]. Based on a similarity characteristic, Adobos was utilized to detect cars and individuals in [13]. The combination of Hear-like feature extraction and Adobos has been utilized to detect the car's rear end using edge features [14]. Knowledge Distillation [15] is a strategy that is referred regarded as a "teacher-student network." The "teacher network" is a more complicated network with exceptional performance and generalization ability. This network is frequently used to teach the "student network," which is less computationally intensive and easier to implement. The knowledge of the "teacher network" is refined into a minimum model by studying the class distribution of the "teacher network." The "student network" then performs similarly to the "teacher network." This strategy drastically decreases the number of procedures and parameters required. However, there are a few flaws as well. This approach can only be used for soft ax loss function classification jobs, which limits its utility (e.g., object detection). Another issue is that the model's concept is far too rigid; resulting in poor performance. The authors presented a method for detecting objects that may be used to support the region proposal network. As a feature extractor, NoCs approved Google Net and Resents. The precision of object detection systems is also improving as neural networks become more complex. However, following the ROI pooling layer, researchers don't pay much attention to merit fusion. The feature fusion module's job is to sort and localize object suggestions. In the Fast/Faster RCNN, the merits fusion module is commonly a multi-layer perceptron. As a result, NoCs investigates the effects of numerous feature fusion modules and discovers that they are just as significant as producing object proposal. The importance of sorting and identifying object proposals is simply highlighted by NoCs. They proposed merit fusion modules, which are extremely complex convolutional neural networks with much more operations and parameters than Faster RCNN [16].Adobos are faster in the testing phase, while SVM are much faster in the learning and training phases. In this research, an application will be designed that uses the detection of objects that use Deep learning algorithm that is programmed in Python language as it is very fast and well used in deep learning. In this research, a system was designed to detect objects using the Yolo algorithm, which is one of the deep learning algorithms, using the Open CV library and using the Python language, and this system is used to help the blind and visually impaired to discover the objects around them. The algorithm was trained on a set of images taken from COCO Dataset, and based on these images, the objects it contains, the algorithm detects the objects that have been trained on it, and this algorithm has proven its high accuracy in detecting objects at a very high speed. The possible applications of this research are Bladafah as it is a system that helps the blind in discovering the objects around them, as this system can be used in large meeting rooms and it knows the people who attended the meeting by training the algorithm on the shapes of people and their names, which is used in residential buildings and give a warning in the event that a stranger enters in addition to other application especially that need high speed in real time

This paper is arranged as follows: section 2 present the Motivation, section 3 present the design and operation of the system. Section 4 present results and discussion and section 5 present the Conclusion.

2. Problem statement

Visually impaired people have difficulty moving safely and independently, which prevents them from participating in routine professional and social activities both inside and outside the home. In addition, as demonstrated in Figure 1, individuals have difficulty identifying principles of the surrounding environment.

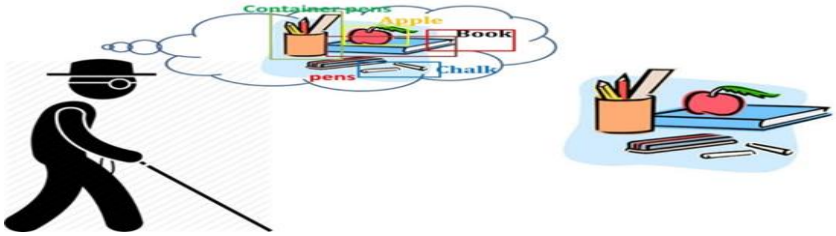

Figure 1: Problem statement

3. The proposed system

The main idea of this system is to make detection and recognition of different objects around the blind in an indoor environment. In this paper the deep learning algorithm being used, whose name is YOLO, and it is a very fast algorithm that is a convolutional neural networks consisting of several layers, each layer have specific job is performed. The idea of YOLO is training CNN (Convolutional Neural Network) on a group of images according to the application used. In this research, CNN was trained on a set of ready-made pictures taken by COCO (Common Object in COntex) and CNN trained on it and after the testing process and to ensure the accuracy of the results. The proposed system is shown in Figure 2.

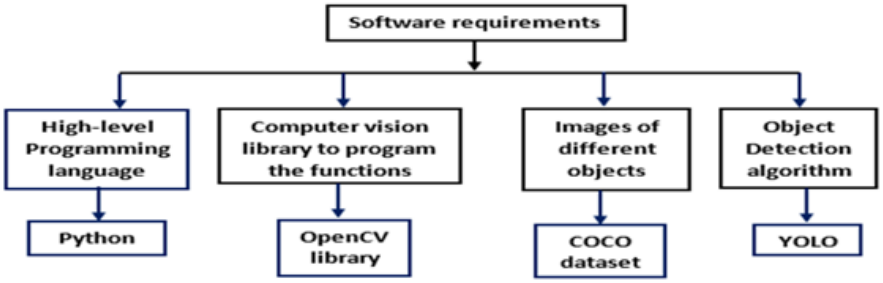

Figure 2: Software stages of the proposed system

3.1 Yolo algorithm

YOLO is a specialized network for objects detection where it performs the process of discovering objects as a single regression problem and takes the input image and passes it through convolutional neural networks and we will get the vector of bounding board class predictions in the output. The principle of the operation depends on the division of the image in to S × S grid cells and each grid cell predicts one object. Each grid cell predicts a fixed number of boundary boxes. Each bounding box represented by five elements $\{b_x, b_y, b_w, b_h, p$, where (b_x, b_y) represent the center of the bounding box, (b_w, b_h) are the box dimensions relative to the image size as shown in the figure below, p represent the probability of the object when object is present in the grid cell, p is set to 1 and p is set to 0 if there is no object in the grid cell [17], see Figure 3. We use a linear activation function for the final layer and the entire following rectified linear layer we use the following activation function:

$$\emptyset\,(x) = \begin{cases} \boldsymbol{X} & \boldsymbol{if\ X > 0} \\ \boldsymbol{0.1X} & \boldsymbol{if\ X < 0} \end{cases} \quad (1)$$

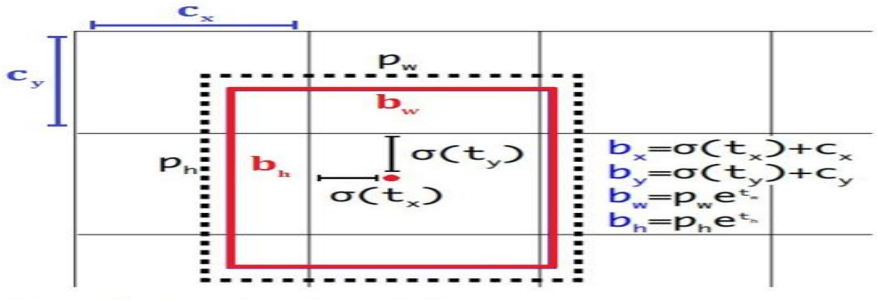

Figure 3: Bounding boxes with dimension priors and location prediction [18]

t_x, t_y, t_w, and t_h, is what the network outputs, C_x,and C_y, and are the top-left coordinates of the grid, pw and ph are anchors dimensions for the box. The operation of YOLO algorithm is illustrated in Figure 4.

$$\mathbf{b_x = \sigma(t_x) + c_x} \quad (2)$$

$$\mathbf{b_y = \sigma_{(t_y)} + C_y} \quad (3)$$

$$\mathbf{b_w = p_w\, e^{t_w}} \quad (4)$$

$$\mathbf{b_h = p_h\, e^{t_h}} \quad (5)$$

$$P_h = \mathbf{e^{tw}} \quad (6)$$

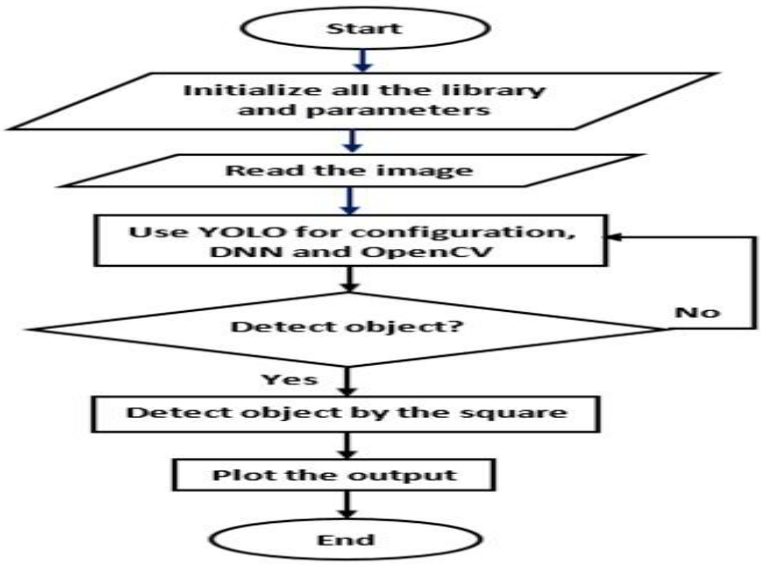

Figure 4: Flowchart of the proposed system

Multiple bounding boxes and sophistication probability for those boxes are predicted simultaneously by a single neural network. YOLO improves detection performance by training on entire photos. This unified model has various advantages over standard object detection approaches.

The YOLO is a very fast but Our regression problem is the detection framework that's wont to be used as a posh pipeline where a replacement image is employed within the testing phase to try the invention of the objects where the differential network works at a rate of 45 frames per second, which suggests that it's possible to process the video accuracy of a vessel of 25 milliseconds from the startup

YOLO algorithm handling images within the conclusion of predictions unlike the algorithms that depend upon the slide window and therefore the proposed area during the training process and therefore the test process you see the YOLO sees the image is all the foremost in order that they are encoding information about the classes.

Fast R-CNN it's one among the great detection methods, but it makes mistakes in background corrections for the objects within the image the rationale is because it doesn't see the entire image, while the Yolo algorithm produces but half the amount of errors within the background compared to YOLO.

When training the YOLO algorithm on real images and conducting the testing process on artworks, the YOLO algorithm excels at detecting objects over other algorithms such as DPM and R-CNN to an outsized degree

The YOLO algorithm uses the features extracted from the input image during the training process and predicts the encompassing boxes. It expects all surrounding boxes through a picture class in sequence. This suggests that the algorithm is fully liable for the image and for all the objects inside the image. This algorithm enables comprehensive image training and real-time speeds while maintaining high average resolution.

3.1.1 Benefits of YOLO algorithm

- Process frames at a pace of 45 frames per second (bigger network) to 150 frames per second (smaller network), which is faster than real-time;
- The network is better at generalizing the image.
- YOLO learns generalize the image better.

YoloV3 has 53 layers and it's neat. Residual network alongside skip connections have improved accuracy & efficiency. Rather than max pooling layers they need used stride convolutional layers which are efficient. Deeper layers increase receptive fields.Predictions are made on 3 features map where they've used FPN style up-sampling to make few features map. It helps in recognizing objects of various scales. On a coco dataset it shows an enormous jump in MAP score and competes with Retina Net and surpasses SSD model. Dark net 53 shows similar results to Resent 101 on Image Net dataset but computational faster. Overall gives better accuracy on coco dataset and a 10x & 100x speed improvement above the previous state of the art.

3.1.2 Problems of YOLO algorithm

- When compared to Faster R CNN, it has a lower recall and a higher localization error.
- Has trouble detecting nearby items because each grid can only suggest two bounding boxes.
- Difficulties detecting small things [19].

Because each grid cell can only include two boxes and have one class, YOLO imposes severe spatial limits on bounding box prognoses. Our model's ability to forecast the number of nearby items is limited by this geographic constraint. This model

has trouble with little things in groupings, such as flocks of birds. Our approach fails to generalize to new or uncommon aspect ratios or configurations since it learns to predict bounding boxes from data. Because the design incorporates several down sampling layers from the input picture, this model also uses fairly coarse characteristics for divining bounding boxes. Finally, the training procedure is based on a loss function that sacrifices detection performance, handling errors as valet in small bounding boxes versus large bounding boxes. A minor error in a large box is usually unnoticeable, whereas a minor error in a small box has a significantly bigger impact on IOU. Wrong localizations are the most common source of error for us.

3.2 Open CV library

It is an open source software library for vision and machine learning and is frequently used in the fields of computer vision. The library contains more than 2500 algorithms which include a set of algorithms from the ancient and modern computer vision [20]. These algorithms can be used to discover faces, identify objects and track the movement of cameras in addition to tracking the movement of the objects themselves. Several images can be linked together to produce a higher solution image. The library is widely used in companies and research groups.

3.3 Coco dataset

In this research, we used a data set called COCO dataset, which is a dataset ready for a number of objects that are frequently used by researchers in the fields of computer vision in general and object detection in particular [21]. Where we trained convolutional neural networks on it and then conduct the testing process. After making sure of the accuracy of the results, the evidence set was expanded and CNN trained on it [22], [23].

Figure 5: Samples of COCO dataset [23]

4. Results and discussion

In this paper Open CV module was used with the pre-trained YOLO model to do objects detection. The module was trained to track objects in the surrounding environments. This model is trained on COCO dataset from Microsoft. It is capable of detecting 80 common objects. The program has been implemented in Python using Anaconda. The experiments have been performed in multiple indoor and outdoor environments with different lights conditions. COCO dataset was used that is available for objects detection, objects segmentation, etc.

In this project, convolutional neuronal networks were trained to perform the process of discovering organisms. CNN has been trained in object detection using 80 objects such as car, book, apple, bear, mobile, bicycle, car etc. After that, the testing process is carried out by inserting images and recognized by the CNN, and the results were excellent in identifying the objects, as ready images and images taken by the regular camera and Raspberry camera were inserted, and this work together was done with an algorithm called YOLO which means that you only look once at the image which is very fast, and this is indicated by its name, because it is very fast. Once you look at the algorithm of the image one time, you will know the existing objects, whether there is one object or several objects in the same image. The results of this work are shown in Figure 6.

Figure 6: The results

5. Conclusion

In this research, a research review using one of the deep learning algorithms specializing in the objects detection of was reviewed and we hope that we have helped people who have vision problems to discover the objects surrounding them. This model is based on training convolutional neural networks, which consist of several layers, each layer is specialized in a specific work, and a COCO dataset has been used. The YOLO algorithm has been trained on these datasets, and we hope in the future that this work will be combined with some developer and find the best results so that they can be applied in various fields and address many problems and facilitate things in various fields.

Author contribution

All authors contributed equally to this work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

[1] Global Data on Visual Impairments 2010. Available online: http://WWW.Who.int/blindness/ GLOBALDATAFINALforweb.pdf (accessed on 23 April 2017).

[2] A. Nada, M. Fakhr and A. Seddik Assistive Infrared Sensor Based Smart Stick for Blind People, in Proceedings of IEEE Technically Sponsored Science and Information Conference London, UK, (2015).

[3] A. Nada, SMashaly, M. Fakhr, and A. Seddik Effective Fast Response Smart Stick for Blind People, in Proceedings of the Second International Conference on Advances in Bio-Informatics and Environmental Engineering – ICABEE, April, (2015).

[4] P. B. L. Meijer, a Modular Synthetic Vision and Navigation. System for the Totally Blind. World Congress Proposals, (2005).

[5] N. Bourbakis and D. Kavraki, Intelligent Assistants for handicapped people's independence: case study, IEEE International Joint Symposia on Intelligence and Systems, Rockville, MD, (1996) 337-344.

[6] N. Bourbakis and P. Kakumanu, Skin-based Face Detection-Extraction and Recognition of Facial Expressions. In Applied Pattern Recognition, .91 (2008).

[7] D. Dakopoulos and N. Bourbakis, Preserving Visual Information in Low Resolution Images During Navigation of Blind, Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environment , Athens, Greece, July, (2008).

[8] N. Bourbakis, Sensing surrounding 3-D space for navigation of the blind - A prototype system featuring vibration arrays and data fusion provides a near real-time feedback, IEEE Engineering in Medicine and Biology Magazine, 27 (2008) 49-55.

[9] V. Santiago Praderas, N. Ortigosa, L. Dunai and G. Peris Fajarnes, Cognitive Aid System for Blind People (CASBliP), Proceedings of XXI Ingehraf-XVII ADM Congress p. 31, June, (2009).

[10] S. Sivaraman and M.M. Trivedi, "Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis," IEEE Trans.Intell. Transport. Syst. 14 (2013) 1773–1795.

[11] L. Shao, X. Zhen, D. Tao and X. Li, Spatio-temporal laplacian pyramid coding for action recognition, IEEE Trans. Cybernet, (2014) 817–827.

[12] W. Liu, X.Z. Wen, B.B. Duan and et al., Rear vehicle detection and tracking for lane change assist, Proceedings of the IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, (2007) 252–257.

[13] T. Liu, N. Zheng, L. Zhao and H. Cheng, Learning based symmetric features selection for vehicle detection, Proceedings of the IEEE Intelligence and Vehicular Symposium, (2005) 124–129.

[14] J. Cui, F. Liu, Z. Li, and Z. Jia, Vehicle localization using a single camera, Proceedings of the IEEE IV, (2010) 871– 876.

[15] Hinton, G.E.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. AR Xiv 2015, are Xiv: 1503.02531.

[16] J.Redom, S.Farhadi, You Only Look Once: Unified, Real-time Objects Detection; 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Electronic ISSN: 1063-6919; Published on 12 December (2016).

[17] Ren, S.; He, K.; Girshick, R.B.; Zhang, X.; Sun, J. Object Detection Networks on Convolutional Feature Maps.

[18] IEEE Trans. Pattern Anal. Mach. Intell. (2017) 1476–1481.

[19] https://arxiv.org/pdf/1506.02640v5

[20] J. Redmon and A. Farha, Yolov3: An incremental improvement. AR Xiv, (2018).

[21] V. Ordonez, G,Kulkarni, and T.Berg, Im2text: Describing images.

[22] T. Lin, M. Maire, S. Belondie,J. Hays,P. Perona,D. Ramanan P. Dollar, and C. L. Zitnick, Microsoft COCO: Common Objects in Context in ECCV,2014.

[23] [Online]. https:// Pjreddie.com/darknet/YOLO/ (Accessed on 14 the July (2018).

[24] C. Arteta, V. Lempitsky, and A. Zisserman. Counting in the wild. In ECCV, (2016).