ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

Real-time object detection of people with visual limitations

[1]BIJOY TAPAN MOHAN NAYAK,
Gandhi Institute of Excellent Technocrats, Bhubaneswar, India
[2]SUKANTA KUMAR JENA,
Vedang Institute of Technology, Khordha, Odisha, India

Abstract: Living with blindness has several difficulties. They encounter numerous difficulties in their daily lives, particularly when moving from one location to another. Accidents happen because of their poor vision. In the area of real-time object detection and recognition for people with visual impairments, a lot of work has previously been done. In layman's terms, we may state that object recognition and object detection both involve identifying the presence of an object in an image or its surroundings. The parameters that have been taken for comparison in this paper's comparative table are dataset, algorithm, and average precision. We have undertaken both theoretical and experimental analysis of the previous works. Additionally, we have identified a research gap in the literature for people with visual impairments, and we have presented our research agenda, which outlines the approach that could have been taken to create a model that is much more practical, i.e., one that can run on low-powered devices like a smart phone.

1. Introduction

In human beings, the eye is the major sensory organ. It helps to visualize the world around us. Without this, one wouldn't be able to find the difference between day and night, blue and black. So, we can assume how difficult it is for the visually challenged to travel from one place to another and to recognize the object around them. According to the fact sheets of WHO (published on 8 Oct 2020)[6]. Globally, 1 billion people have a problem with vision impairment. It includes all types of impairment like trachoma, glaucoma, uncorrected refractive index, cataract, age-related macular degeneration, corneal opacity, diabetic retinopathy. So, for making their life a little bit easy we can provide them with the vision of a computer. We can provide vision to visually challenged people by object detection and recognition and by informing them about their Surroundings using some auditory device like headphones etc.

Object recognition is a kind of simple process for human beings but for computers it is not that easy task as it consists of a step-by-step process of recognizing, identifying, and locating the objects with input with a given degree of precision. Recognition basically consists of classification and detection. Objects can be divided into their respective classes by performing three steps - feature extraction, localization, and classification on the objects. In classification, the algorithm recognizes the class of the object with a degree of confidence. After classification, we know that the particular class of the objects from which this object belongs. Now, in detection, we put a bounding box around the object in the picture.

The main objective of this work is to present a comprehensive and comparative analysis of the work that has been done in the field of object detection for visually challenged people. We will present here the comparative analysis of the algorithms that have been used in existing systems. Basically, we divide the object detection algorithms into two categories, first, one category is region-based object detection algorithms and the second category is regression-based object detection algorithms. The main advantage of regression-based

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

algorithms over region-based algorithms is that regression-based algorithms work faster in comparison to region-based algorithms. Here, we will present the gap in research work on the basis of the papers that we have read and we will propose our system to fill that research gap.

After analysis of the research papers related to object detection and identification for visually challenged people, we came to know that lots of work have already been done in this field but we didn't find any model suitable for low computation devices like a mobile phone without any dependency(like an external server for GPU). The models which are present in the research papers either require additional hardware or server connectivity because they are using algorithms that require faster processing devices like GPU. They should have used the algorithms which can easily work on a low computation device.

2. Overview

Here, we have presented a brief overview of the procedure and internal working of the object detection and identification systems. Figure 1 shows the various steps which are involved in the process of object detection and identification.

In order to detect and identify the objects, the first step is the collection of datasets for the training and testing of the model. A custom dataset can be created by manually capturing and labeling the images. A number of datasets for object detection are also available over the internet. Few popular available datasets are MS COCO (Provided by Microsoft), ImageNet, and Pascal VOC.

After the successful creation of the dataset, the next step is to divide the dataset into the training dataset and the testing dataset. After the division of the dataset into training and testing dataset, the next step is to choose the object detection algorithm. All the existing object detection algorithms can be classified into two categories, the first one is a region-based object detection algorithm and the second one is a regression-based object detection algorithm. According to the system requirement, algorithms can be chosen and the system can be trained on the chosen algorithm. In the next step, object detection is performed on the given image. If the system detects any object using DNN, identification is performed over the object.

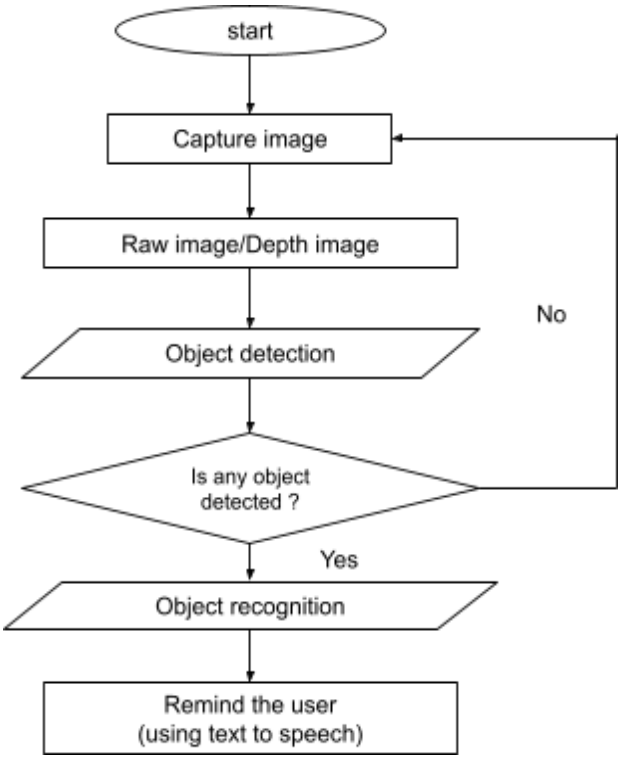

Figure 1:: Structural outline of object detection and identification unit.

3. Literature Review

Review Methodology

A platform like google scholar has been used for searching the papers related to our work. The papers have been searched on the basis of a few keywords like object detection in real-time, object recognition for visually challenged people using smartphones, and models for visually challenged people. After getting lots of research papers we have selected a few research papers on the basis of the abstract of research papers. We have read the introduction and conclusion of the selected research papers then we have excluded some research papers and finally selected the research papers which were closely related to our work.

Related Work

A lot of research work has already been done for object detection and identification. Existing works show how the problem of visually challenged people can be solved. In existing works the devices which are used as

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

visual substitution takes the input from the surroundings of the VI people and draws out the information about the objects which are present in their surroundings after this, the system notifies visually challenged people about their surroundings by using some auditory device. Existing work is comparable on different-2 parameters but here we are performing the comparison of the previous work on the basis of algorithms and datasets that have been used in the existing systems. All the researchers have used either a region-based algorithm or a regression-based algorithm in their work. Region-based algorithms are popular for their accuracy and regression-based algorithms are popular for their speed. For real-time object detection, most of the researchers have used regression-based algorithms. In short, regression-based algorithms like YOLO, MobileNet SSD are more popular for object detection nowadays. Here, after compiling a number of research papers on the basis of some criteria (selection criteria have been discussed in review methodology), we have selected 7 research papers for presenting the comparative analysis of the existing work. Out of these selected research papers, 4 are using the regression-based object detection algorithms, 2 are using region-based object detection algorithms and 1 is using both region-based and regression-based object detection algorithms.

Some of the existing research work related to object detection and identification for visually challenged people have been discussed here –

Prateek Agrawal et al.[1] has proposed a system for bank cheque verification. This system verifies the bank cheque using the following information - cheque number, bank account number, bank branch code, legal as well as the courtesy amount, and signature. They have used the IDRBT cheque dataset and deep learning-based CNN with high accuracy of 99.14% for handwritten digit recognition. In this system, for the signature verification, they have used SIFT feature extractor and SVM (Support Vector Machine) as a classifier with high accuracy of 98.10%.

Sandipan Chowdhury et al.[5] has proposed a method for object detection through a webcam. In this method, they have used a combination of Fast R-CNN (Region-based Convolutional Neural Network) and Region Proposal Network (RPN), where high-quality region proposals are generated by training the RPN end to end, which are in turn is used by Fast-RCNN for detection. These two modules combine to generate an object detection system called Faster R-CNN. Faster R-CNN algorithms can detect objects in real-time with very high speed still there are few algorithms that are faster than Faster R-CNN.

Cheng Qian et al. [10] has developed an indoor wayfinding system. In this system, the YOLOv2 algorithm has been used for the detection of indoor objects like doors, door handles, etc. The advantage of using the YOLO algorithm is its high speed. In this system, a visually challenged person is connected with a portable camera, with Bluetooth earpiece and GPU. This system contains three main components, a deep neural network, a camera, and an auditory device through which the subject can get to know about the objects around him. Cheng Qian et al. has used a convolutional neural network (CNN) in this model. The ConvNet which is used in this model has 22 layers and that's why this model is perfect for identifying items as fast as the input images are classified into the matrix in the primary layer where bounding box offers are upraised. The requirement of extra hardware devices, GPU, and the stereo camera makes this model costly as well as an extra burden for visually challenged people.

Bor - Shing Lin et al.[9] have also used the YOLO algorithm in their system. This model is worked on a smartphone and a server. This model is worked in two different modes, online and offline mode, The online mode was "stable" and the offline mode was "fast". In both modes, the model works such that the smartphone extracts the features from the surroundings and the server provides information about the direction and distance of the objects. Although, in the "stable" mode, the server uses the Fast R-CNN algorithm whereas in the "fast" mode it works on the YOLO algorithm. The Fast R-CNN has been used here, for object identification and for roughly estimating the distance and position of the object, Fast R-CNN makes the system more accurate but a little slower. The YOLO algorithm is used herein Fast mode, as YOLO looks only once to take out both information so this algorithm has been used here to make the system faster. In this system object detection and identification using smartphones requires a dedicated server on which YOLO and Fast-RCNN models can run. Its dependency on the internet is also a negative point in this system, as the system performance is dependent on the internet, it would not be able to detect object accurately and in a fast way, if there would be any internet issue occurs and it will be bad for VI people because they can't depend on the internet.

F Particke et al.[11] has developed a system for real-time object detection and localization by the use of a smartphone platform. In this system, F Particke et al.

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

have used neural networks for object detection and localization in real-time. Hence, detection instructions of a DNN are connected with the depth data from a Depth Sensor(RGB-D camera), and that RGB-D camera is staged on a smartphone platform. In this system, the YOLOv2 algorithm was the choice of researchers as an object detection algorithm. In this system, the localization can be further advanced in future works by correspondingly considering spatial information in the clustering system. By using the X-means algorithm instead of the K-means algorithm object detection and localization can be furthermore improved. The problem with the YOLO algorithm is that it requires a GPU to run and it can detect objects only in the range of 2m to 5m so, there is also the possibility of improvement.

Hanen Jabnoun et al.[7] have done object detection for visually challenged people in video scenes. In this system, the feature extraction and the RANSAC algorithm are used to fit the model.SIFT feature extraction algorithm is the scale and angle invariant so its accuracy is decent. But the used in this system is very small which contains only 8 classes. This system will not perform well in real-time object detection because the algorithms which are used in this system are slow for real-time object detection.

Chugai Yi et al.[14] has developed a system for searching objects for assisting blind people. This model contains a wearable camera and different – different fixed cameras. The visually challenged person is connected with a wearable camera and this camera is bound with a computer system, the subject can make a request for searching any object through the voice command and then wears that camera for getting the object location and when the system identifies the requested object a voice message would be generated. In this system, the SURF algorithm has been used for feature extraction and the SURF algorithm provides high accuracy in object detection and identification but in future work, this system can be made more efficient by using some fast object detection algorithm so that in real-time the system can work in a better way. As we know fast object detection and identification is required when we detect objects in real-time.

Object Detection and Identification is the area that was targeted by researchers. Here, Table 1 presents the comparative analysis of the research work in Object detection and identification for visually challenged people. In figure 2 we have presented the general architecture of object detection and recognition using DNN.

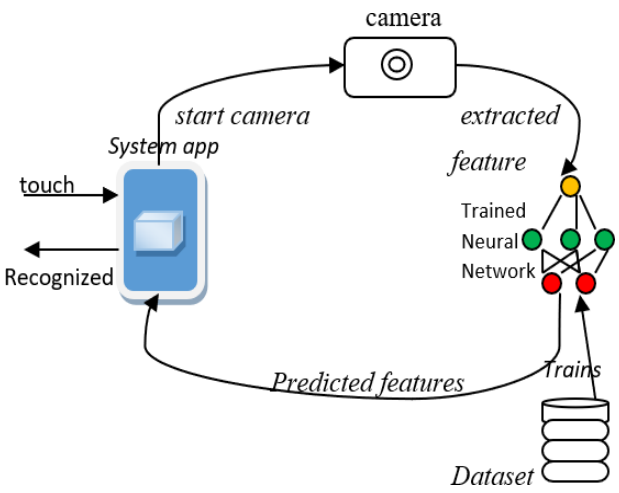

Figure 2:: Common architecture for object detection using neural network

The architecture shows that when the system gets a command to detect an object it takes the image data as input from its surrounding in real-time using the system camera and provides that input data to a trained neural network which predicts the feature in the given image and shows the result on the system and if the object is recognized it notifies to the user.

Experimental Overview

Every algorithm has its own advantages and disadvantages. In table 2, we are presenting the experimental comparison of the YOLO algorithm with its different versions and with EfficientDet on the two different datasets that are MS COCO and DOTA datasets. In table 2, we have shown the comparative analysis of YOLOv2, YOLOv3, YOLOv4, and EfficientDet on 4 different parameters. These parameters are – FPS, mAP, BFLOPs (Floating point operations in billions), and model size.

As we know YOLO and EfficientDet both are regression-based algorithms and regression-based algorithms are very much popular for their high speed. We can see from the provided analytics, In the starting version of YOLO i.e. YOLOv2, speed is very high but the value of mAP is 17.6 which is very low whereas in the newer version of YOLO speed is high and accuracy is also better in comparison to an older version. In YOLOv4 speed is 18.5 FPS and the mAP is 55.4 which is very good. In the earlier version of the YOLO, the number of the floating-point operations (BFLOPs) that are performed in one second are very less whereas in the

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

later versions of YOLO the number of the floating-point operations (BFLOPs) that are performed in one second is very high and this is the reason for the higher accuracy of the later versions of the YOLO in comparison to earlier versions of the YOLO. Due to less number of floating-point operations per second in earlier versions of YOLO, these versions of YOLO are able to process more frames in one second (FPS). Model size has also increased the newer version of YOLO. If we compare EfficientDet with YOLO, both can compete with each other in terms of accuracy but in terms of speed, YOLO is far better than EfficientDet.

Table 1
Comparison of previous work on Object Detection and Identification based on different parameters and datasets (*AP- Average Precision, *RR- Recognition Rate)

Author	Domain	Dataset	Algorithm	Metric Name	Metric Value
Prateek Agrawal et al. 2020	Bank Cheque Verification using Deep Learning	IDRBT Cheque dataset	SIFT & SVM	AP	98.10
Sandipan Chaowdhury et al. 2019	Real-Time Object Detection using Deep Learning: A Webcam Based Approach	Custom Dataset	Faster RCNN	AP	95
Chandan G et al. 2018	Real-Time Object Detection and Tracking Using Deep Learning and OpenCV	COCO	Mobile Net	AP	99
Berriel et al. 2018	Crosswalk Detection for VI Pedestrian	GSV dataset, IARA dataset	VGGnet	AP	96.51 on GSV, 92.04 on IARA
Cheng Qian et al. 2017	Door Knob Detection for Visually Challenged People with Feedback in Real-Time	Custom Dataset	YOLOv2	AP	80
Bor-Shing Lin et al. 2017	Smartphone-based assistive system for visually challenged people	ImageNet	Faster RCNN & YOLO	AP, RR	57-approx, 60
F Particke et al. 2017	Deep Learning for Real-Time capable object detection and localization on mobile platform	ImageNet	YOLOv2	AP	88
Joseph Redmon et al. 2016	Real-time Object Detection with YOLO	Pascal VOC 2007, ImageNet	YOLO	AP	59.2 on Pascal VOC, 88 on ImageNet
Chugai Yi et al. 2013	Finding objects for assisting blind people	Custom Dataset	SURF and SIFT	AP	69 on SURF, 73 on SIFT

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

Table 2
FPS (Frames Per Second), MAP (Mean Average Precision), and BFLOPs score on Various Datasets.

Model	Model Size	Dataset	FPS	mAP	BFLOPs
YOLOv2	**202.3M**	**DOTA**	**58.3**	**17.6**	**44.417**
YOLOv3	**245.78M**	**DOTA**	**14.7**	**55.8**	**101.784**
YOLOv4	**246.3M**	**MS COCO**	**18.5**	**55.4**	**128.459**
EfficientDet	**234.56M**	**MS COCO**	**6.5**	**55.1**	**40.223**

4. Findings

After analyzing the existing work of object detection and identification for visually challenged people, we have presented here the detailed comparative analysis of the work in the form of table 1. Most of the existing systems have used regression-based algorithms like YOLO and MobileNet SSD for object detection and identification. Obviously, the reason is the fast speed of YOLO and MobileNet SSD over the region-based object detection algorithms. In terms of accuracy, region-based object detection algorithms are better than regression-based object detection algorithms but in real-time object detection systems, priority is always speed. Region-based algorithms are used where accuracy is a priority.

Problems In Existing Method

As in the existing system, most of the real-time object detection systems use the YOLO object detection algorithm for detecting and identifying objects due to its fast detection and identification speed, very few systems use R-CNN or Faster R-CNN. But the problem with the YOLO is less accurate in comparison to Fast R-CNN or other Region-based CNN. YOLO can detect objects only in the range of 2-5m but what about the objects which are not present in that range. Identifying the object outside the 5m requires some other model. Extra hardware device requirement is also a negative point for visually challenged people in existing systems because all the time they can't carry the extra burden with them. Comparative analysis of the most popular object detection algorithm which is most suitable for real-time object detection and identification has been provided by Ugur Alganci et al. [2] on the DOTA image. The work of Ugur Alganci et al. shows that almost in every case SSD(Single Shot Detector) algorithm has the least recall time and the precision of the SSD algorithm is also high, the precision of Faster RCNN and YOLO is also high but recall time is not as good as SSD.

Comparison between Region-Based Algorithms and Regression-Based Algorithms:

Regression-based object detection algorithms are faster than region-based object detection algorithms because region-based object detection algorithms detect objects in three phases whereas regression-based algorithms detect objects in a single phase. In region-based algorithms, firstly it generates the region proposal, the second phase is feature extraction and the third phase is classification and object detection. The region-based object detection algorithm follows the sliding window technique for generating the region proposal then it uses feature extraction algorithms (like SURF, SIFT, etc.) for feature extraction and in the last phase, it uses a classifier for classification. Regression-based object detection algorithms divide the images into grids and provide 0 or 1 to each grid

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

according to the object presence and absence, that is how regression-based algorithms detect objects in a single phase. As the name of regression-based algorithms also shows these algorithms work in single-phase like YOLO stands for You Only Look Once and SSD stands for Single Shot Detector. In terms of accuracy, region-based object detection algorithms are better than regression-based object detection algorithms. We use region-based object detection algorithms where accuracy is a priority and we use regression-based object detection algorithms where speed is a priority.

5. Research Agenda

We should develop such a system that doesn't require any extra device for object recognition for visually challenged people. Our proposed system requires just a single device i.e. smartphone, which is easily available to all in this era. We need to use such an algorithm that requires fewer computational devices and can detect and identify objects in real-time with accuracy as well as fast speed. SSD (Single Shot Detector) and based algorithms would be more appropriate for object detection and identification in real-time with low computational devices. Faster RCNN and SSD have better accuracy in comparison to YOLO, YOLO gets more credit when we prioritize speed in the system in comparison to accuracy. In DNN, SSD with Mobile Nets performs more efficiently. SSD and Mobile Net detects the object with accuracy and with fast speed also.

SSD (Single Shot Detector) can detect multiple objects in an image in a single shot, on the other hand, region-based neural network algorithms take three steps for detecting objects, one for generating region proposals, the second for feature extraction, and the third one for classification for detecting objects of each proposal. The experimental results presented by Ugur Alganci et al. [2] show that the Average Precision (AP) of this algorithm to detect different classes as a car, person, and chair is 99.76%, 97.76%, and 71.07%, respectively. This raises the correctness of object detection at a processing speed that is needed for real-time detection and thus fulfills the need for regular monitoring indoor and outdoor.

6. Conclusion

In this paper, we have presented the comparative analysis of the existing work which has been done in the field of object detection for visually challenged people. This analysis shows that existing systems are not that appropriate and useful for visually challenged people. Either existing systems are not useful in practical scenarios or they require some type of dependency. Most of the systems use extra hardware devices which becomes an extra burden for visually challenged people. There is a need for such a system which couldn't feel like an extra burden to them and also be a part of their life. Our proposed system can fulfill this need of visually challenged people. Our proposed system is an Android Application for visually challenged people for object detection and identification which will help them by informing them about their surroundings. It will help them by performing indoor as well as outdoor object detection and identification. After identification of the object system will inform the visually challenged people by generating a voice.

7. References

[1] Agrawal, P., Chaudhary, D., Madaan, V., Zabrovskiy, A., Prodan, R., Kimovski, D., & Timmerer, C. (2020). Automated bank cheque verification using image processing and deep learning methods. *Multimedia Tools and Applications*, 1-32.

[2] Alganci, U., Soydas, M., & Sertel, E. (2020). Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. *Remote Sensing*, 12(3), 458.

[3] Berriel, R. F., Rossi, F. S., de Souza, A. F., & Oliveira-Santos, T. (2017). Automatic large-scale data acquisition via crowdsourcing for crosswalk classification: A deep learning approach. Computers & Graphics, 68, 32-42.

[4] Chandan, G., Jain, A., & Jain, H. (2018, July). Real time object detection and tracking using Deep Learning and OpenCV. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1305-1308). IEEE.

[5] Chowdhury, S., & Sinha, P. Real Time Object Detection using Deep Learning: A Webcam Based Approach.

[6] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

ISSN: 2278-4632
Vol-10 Issue-09 No.03 September 2020

[7] Jabnoun, H., Benzarti, F., & Amiri, H. (2015, December). Object detection and identification for blind people in video scenes. In *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 363-367). IEEE.

[8] Ju, M., Luo, J., Zhang, P., He, M., & Luo, H. (2019). A simple and efficient network for small target detection. *IEEE Access*, 7, 85771-85781.

[9] Lin, B. S., Lee, C. C., & Chiang, P. Y. (2017). Simple smartphone-based guiding system for visually impaired people. *Sensors*, 17(6), 1371.

[10] Niu, L., Qian, C., Rizzo, J. R., Hudson, T., Li, Z., Enright, S., ... & Fang, Y. (2017). A wearable assistive technology for the visually impaired with door knob detection and real-time feedback for hand-to-handle manipulation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 1500-1508).

[11] Particke, F., Kolbenschlag, R., Hiller, M., Patiño-Studencki, L., & Thielecke, J. (2017). Deep learning for real-time capable object detection and localization on mobile platforms. *IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK*.

[12] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[13] Saranya, N., Nandinipriya, M., & Priya, U. (2018). Real time object detection for blind people. *International Journal of Advance Research in Science and Engineering, 7(1),* 306-316.

[14] Yi, C., Flores, R. W., Chincha, R., & Tian, Y. (2013). Finding objects for assisting blind people. *Network Modeling Analysis in Health Informatics and Bioinformatics, 2(2),* 71-79.