Received: 11 October 2022 | Revised: 1 March 2023 | Accepted: 7 March 2023 | Published online: 28 March 2023

REVIEW

Artificial Intelligence and Applications
2023, Vol. 1(3) 139–144
DOI: 10.47852/bonviewAIA3202456

BON VIEW PUBLISHING

Advances in Real-Time Object Detection and Information Retrieval: A Review

Shital N. Katkade[1,*], Vandana C. Bagal[2], Ramesh R. Manza[1] and Pravin L. Yannawar[1]

[1]*Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, India*

[2]*Karmaveer Kakasaheb Wagh Institute of Engineering Education and Research, India*

Abstract: Visually disabled people's day–night lives are delicate, they are facing numerous problems when traveling from one position to another, and they are more likely to be involved in an accident as a result of their lack of vision. The motive of this review paper is to explore colorful ways used by other experimenters worldwide for persons with vision loss to fulfill their full eventuality. The system alerts visually impaired individuals about their surroundings by employing some sort of audio device, extracting information about the objects that are present in their surroundings using the devices now in use as visual substitutes. Utmost results handed by experimenters bear fresh tackle, which adds to the burden for visually disabled people in the world. A system is required that will help them in their day–night lives and become part of their life and will not feel like a burden. The dataset was used by experimenters for object detection, COCO (handed by Microsoft), Pascal VOC, ImageNet, etc., and this dataset is publicly available on the internet.

Keywords: deep learning, raspberry pi, object detection, region based, regression based, neural network, computer vision

1. Introduction

This day, the life of visually disabled people is relatively grueling; they face many complications in their daily lifestyle. The eye is the most important sense organ in humans. It aids in the visualization of the world around us. Without it, it would be impossible to tell when it was day or night, blue or black. As a result, we can imagine how delicate it is for the visually disabled to travel from one position to another and to identify the objects in their surroundings. Object recognition is a fairly straightforward process for humans, but it is not so simple for computers, as it involves a step-by-step method for detecting, locating, and finding specifics. Recognition is made up of two parts: classification and detection. Feature extraction, localization, and classification are three procedures that can be used to separate objects into their applicable classes. As a result, real-time object finding has come a critical issue for the mobilization of human tasks. Real-time object discovery refers to when something happens in real time and can be inferred rapidly while keeping a high position of precision. For object detection and identification, a lot of research has previously been completed. Being works demonstrate a solution to the issue of blind people. The main purpose of this exploration is for the eyeless points at expanding possibilities to people with vision loss to realize their full potential, most system includes redundant tackle bias, which adds an unnecessary burden on those who are blind or visually impaired. There is a need for such a system that may not only be a part of their lives but also not feel like a redundant burden to them. The major goal of this review effort is to provide a thorough and comparative examination of the research that has been done in the area of object detection for people who are visually impaired. Then, we will give a comparative examination of the algorithms that were applied to create the systems. The experimenter classifies the object detection algorithms into two basic groups: region-based object detection algorithms and regression-based object detection algorithms. Regression-based algorithms have a significant performance advantage over region-based algorithms due to their faster processing times.

*Corresponding author: Shital N. Katkade, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, India. Email: shitalkatkade25@gmail.com

2. Literature Review

2.1. Real-time object detection and identification for visually challenged people using mobile platforms (Joshi et al., 2021)

The live works have been subjected to theoretical and experimental study by Joshi et al. (2021) in the form of a relative table, where the parameters used for comparison are dataset, algorithm, and average perfection. They distinguish between two different types of object detection algorithms: a region-based object detection technique and a regression-based object detection technique. Regression-based algorithms are well-liked for their speed, while region-based algorithms are well-liked for their precision. The experimental comparison of the You Only Look Once (YOLO) algorithm with EfficientDet and its various performances on the two distinct datasets, MS COCO and DOTA, is described in this work. After evaluating the current state of object detection and identification for visually impaired individuals, they established that region-based algorithms are superior to regression-based algorithms in terms of accuracy, but speed is always given priority in real-time object detection systems. Where accuracy is valued highly, region-based algorithms are employed. The experimenter creates a system so that visually impaired persons can find objects without the use of any additional devices. Their suggested solution only needs a single gadget, a smartphone, which is widely accessible to everyone.

2.2. Deep learning in multi-object detection and tracking: State of the art (Pal et al., 2021)

Pal et al. (2021) offered a thorough analysis of a number of deep learning (DL)-based models for the tasks of general object detection, specific object detection, and object tracking, taking both the detection and tracking into account separately as well as in combination. They conducted comparative studies between various detectors, such as two-stage detectors and one-stage detectors. In two-stage detectors, object regions are first proposed using a region proposal method before features are used for classification and drawing bounding boxes around objects, whereas in one-stage detectors, bounding boxes are directly predicted in videos or images without the use of a region proposal method on various types of data, such as PASCAL VOC, MS COCO, ImageNet, and MOT. Performance metrics called mean average precision (mAP) were applied. Tables 1 and 2 give a comparative analysis of different detectors on different data.

Table 1
Object detection result of various detectors on MS COCO dataset

Method	Data	Backbone	(%)
Fast RCNN	train	VGG-16	19.7
Faster RCNN	trainval	VGG-16	21.9
R-FCN	trainval	VGG-16	22.6
CoupleNet	trainval	ResNet-101	34.4
Faster RCNN+++	trainval	ResNet-101-C4	34.9
Faster RCNN w FPN	trainval35k	ResNet-101-FPN	36.2
Deformable R-FCN	trainval	Alignmed-inception-ResNet	37.5
wnQ,-ted	trainval	ResNet-101	40.8
Mask RCNN	trainval35k	ResNetXT-101	39.8
DCNv2+Faster RCNN	train118k	ResNet-101	44.8
YOLOv2	trainval35k	DarkNet-53	33.0
YOLOv3	trainval35k	DarkNet-19	21.6
DSSD321	trainval35k	ResNet-101	28.0
SSD513	trainval35k	ResNet-101	31.2
DSSD513	trainval35k	ResNet-101	33.2
RetinaNet500	trainval35k	ResNet-101	34.4
RetinaNet800	trainval35k	ResNet-101-FPN	39.1
M2Det512	trainval35k	ResNet-101	38.8
M2Det800	trainval35k	VGGI6	41.0
RefineDet32o+	trainval35k	ResNet-101	38.6
RefineDet5 I2+	trainval35k	ResNet-101	41.8
FPN	trainval35k	ResNet-101	39.8
NAS-FPN	trainval35k	RetinaNet	40.5
NAS-FPN	trainval35k	AmoebaNet	48.0
Granulated CNN	trainval35k	ResNet-101	32.2

Table 2
Object detection result of various detectors on PASCAL VOC dataset

Method	Training data	Test data	Region proposal	Backbone	(%)
RCNN	VOC 07	VOC 07	SS	AlexNet	58.5
RCNN	VOC 07	VOC 07	SS	VGG16	66.0
Fast RCNN	VOC 07+ VOC 12	VOC 07	SS	VGG16	66.9
YOLO + Fast RCNN	VOC 07 + VOC 12	VOC 12	SS	VGG16	70.7
YOLOv2	VOC 07 + VOC 12 +MS COCO	VOC 12	–	DarkNet-53	78.2
Fast RCNN	VOC 07 + VOC 12	VOC 12	SS	VGG16	68.4
Faster RCNN	VOC 07 + VOC 12	VOC 12	RPN	VGG16	70.4
Faster RCNN	VOC 07 + VOC 12 + MS COCO	VOC 12	RPN	VGG16	75.9
YOLO +Fast RCNN	VOC 07 + VOC 12	VOC 12	RPN	VGG16	70.7
YOLOv2	VOC 07 + VOC 12 + MS COCO	VOC 12	–	DarkNet-53	78.2
SSD300	VOC 07 + VOC 12 + MS COCO	VOC 12	–	ResNet10l	79.3
SSD512	VOC 07 + VOC 12 + MS COCO	VOC 12	–	ResNet10l	82.2
R-FCN	VOC 07 + VOC 12 + MS COCO	VOC 12	RPN	ResNet101	85.0
G-RCNN	VOC 07 + VOC 12	VOC 12	FRPN	G-AlexNet	80.9

2.3. Real-time object detection for visually impaired using OpenCV (Balaji et al., 2020)

In this paper, Balaji et al. (2020) created a model to identify items at a specific distance for vision-impaired individuals and other commercial reasons. Older methods for object detection required a lot of training data, which added to the task's complexity and length. This system requires importing a pre-trained model from the DL framework caffemodel. For object detection, the caffemodel framework offers a MobileNet model that imports pre-trained objects. The deep neural network (DNN) algorithm is used by the model to recognize and classify items. The system has been taught to accept items with a confidentiality level of at least 50% required for object detection. Better simultaneous image detection is made possible because of this. A portable device with the necessary coding, camera, and sensors for object detection was needed for the system. The laptop's web camera or any personal device can be used to view the output of the primary working model, which is run in Python code. They employed the MobileNet–single shot detector (SSD) model, a pre-trained MobileNet–SSD detection network implementation in Caffe. A quicker and more effective DL-based object detection approach is produced by combining the MobileNet architecture with a SSD framework. The quickest detectors provide the best accuracy when using SSD and MobileNet. The proposed system has an object detection accuracy of up to 95%. This model has various drawbacks, such as MobileNet–SSD being less accurate with smaller items and finding things when there are backdrop distractions or if the images are blurry. In the future, MobileNet architecture should be enhanced to more accurately recognize smaller objects and objects with lower confidence levels.

2.4. Real-time object detection for visually challenged people (Vaidya et al., 2020)

The system suggested in this research aids visually impaired people in reducing their mobility and productivity in carrying out daily duties. As soon as the application launches, it begins capturing a live video stream an input from the camera. Objects are detected in the camera's field of view along with their approximate location, which is communicated to users via audio output, according to Vaidya et al.'s (2020) implementation. The YOLO method, which is best suited for real-time object detection applications, is used in this system, and what distinguishes YOLO from other algorithms is its excellent processing speed, where it can process 45 frames per second, and a small YOLO dataset is used to run the object detection algorithm in the Android application. This model achieved a maximum accuracy of 85.5% for mobile phones and 89% for web applications; although in most cases the model performs accurately in detecting seven different objects, it may not perform well in cases where the object is too close to the camera or is not part of the trained dataset.

2.5. Object detection and distance estimation tool for blind people using convolutional methods with stereovision (Bastomi et al., 2019)

In this study, Bastomi et al. (2019) created a tool to obtain information about objects. This system can also use a camera attached to the glasses to estimate the distance of detected objects to facilitate the use of blind users. The primary sensor is a camera, which functions similarly to human eyes and produces real-time video as visual data. The next step in this research is that the input image is acquired by a web camera connected via a USB port. It was processed with OpenCV, then set to 640×320 pixels, and then processed with a convolutional neural network (CNN). CNN is used in the image processing, where the weight obtained from the image is classified according to the mass of the image in the material. CNN is usually divided into two layers, a feature extraction layer and a classification layer. The feature extraction layer has a convolution layer and a max-pooling layer; the convolution layer is used to reduce the pixel size of the image, while the maximization layer is used to maximize the sharpness so that the image is clearly visible. The basic equipment is a mini PC, a camera, and a battery. The result of this study is that the performance of the CNN system in detecting objects in the form of cars, tables, chairs, bicycles, people, and motorcycles from different directions has its own characteristics. There are still inappropriate recognition results. According to the results, the success of observations is quite high, 93.33%, and the stereo video measurement results have a large error of about 6.1% when the object is directly between the two cameras. Stereo vision is weak, if the detected object is not directly in front of and between the two cameras, the stereovision measurement cannot measure correctly. Stereovision has limited distance measurement, measuring a distance of only 50 cm to about 300 cm.

2.6. Object detection system for the blind with voice guidance (Karmarkar & Honmane 2021)

A blind-friendly object detection system has been created by Karmarkar & Honmane (2021) using DL technology. The position of things is also explained to visually impaired people using a voice-guided technique. To make it simpler for the blind to learn about objects, researchers employed the YOLO technique in the object identification DL model and synthesized a vocal announcement using text-to-speech (TTS). She used the COCO dataset, which has 330 images and 80 labels, to train an object detector model. As a result, they were able to create an Android application for object detection that displays the name of the object and its accuracy as a percentage. The class label and estimated portion of the recognized object are combined as text, and the Google TTS Application Programming Interface (API) may be used to provide audio feedback. The accuracy of this research can be improved with further investigation. Additionally, the current system runs on the Android operating system, which may be improved further to work with any convenience devices.

2.7. Object detection and identification for blind people in video scene (Jabnoun et al., 2015)

The aim of this work is to provide a proposed system that restores an important role of the visual system: object detection. This method is based on local feature extraction. The simulation results of object detection using the Scale-Invariant Feature Transform (SIFT) method and key point matching showed good accuracy. The proposed visual compensation system is based on the recognition of objects around the blind person. We offer a system that detects and recognizes 2D in the video. This system should find invariant features of objects to change perspective, provide recognition, and reduce detection complexity. Jabnoun et al. (2015) propose a method based on video key extraction and matching. A comparison between the query frame and the database objects is performed to identify the object in each frame. For each detected object, an audio file containing information about it is activated. Thus, object detection and recognition are processed simultaneously. They create a bag of object key points using SIFT and compute a local difference map between frames. This algorithm emphasizes the use of SIFT to identify and match key points to ensure object detection. The key point comparison problem is to figure out how the key points of different frames are matched to the key points of the target objects. The researchers found a large percentage of the detected objects, but they also tried to figure out why some of them failed.

2.8. Real-time object detection and multilingual speech synthesis (Stephen et al., 2019)

In this paper, Stephen et al. (2019) present a way to produce many voices from a single model by fusing real-time deep-learning picture recognition with multilingual neural TTS synthesis. They demonstrate that the existing monolingual approach has been enhanced by using a single-model neurotext-to-speech synthesis system. The model is separated into five subtypes: image acquisition unit, classification unit, TTS, prediction point domain, and output unit. They employed a transfer learning technique on the Inception v3 model, one of the most well-known DL models for computer vision tasks. They learned about the exception model through transfer learning. By pre-training a DNN, transfer learning is often used in conjunction with DL and has been used effectively in a variety of computer vision tasks. They used it to extract common image features such as edges; this information was then used to classify the images or live video stream and then fed back to the model after classification, and finally the conversations started after the TTS operation was done with the word associated with the picture. As a result, they showed how to use a single model to recognize real-life objects that were spoken in multiple languages. This technique can be applied to more complex problems such as scene image segmentation and measuring the distance between objects using distance estimation. This could open the way for major real-time improvements and applications in areas such as autonomous multilingual learning, human–machine communication, voice-based object-to-object tracking, and even assisting the visually impaired.

2.9. DL-based object detection and recognition framework for the visually impaired (Bhole & Dhok, 2020)

In this paper, Bhole & Dhok (2020) provide a unique framework to help the blind and visually impaired move independently and be aware of their surroundings by assisting them in object detection and recognition. The research uses transfer learning on a single shot detection mechanism for object identification and categorization, followed, if identified using the Inception v3 model, by the recognition of human faces and money notes. With the addition of a new class, the SSD detector was trained using a modified version of the PASCAL VOC 2007 dataset, allowing for the identification of cash as well. The framework's output can ultimately be delivered to the visually impaired person in audio format. The standalone SSD detector's enhanced accuracy class mean accuracy and precision 10 (mAP) score was 67.8%, and the Inception v3 model test accuracy was 92.5% and 90.2% for the person and currency detection.

2.10. CNN-based auto assistance system as a boon for directing visually impaired person (Shah et al., 2019)

The goal of Shah et al. (2019) is to create an object detection system that will allow persons who are fully blind to perform their everyday tasks on their own. They contrast various approaches to object detection, including CNN and the Haar cascade. An approach to DL that can be used for object detection is the CNN, whereas the Haar cascade has been trained to recognize a variety of items, specifically for face detection but also used for object detection. The 2300 images in the custom dataset, which contains three separate classifications, were produced by researchers. They used the COCO 2017 dataset to train a CNN model, and as a result, the Haar cascade algorithm is the fastest one, and CNN provided greater object detection accuracy in real time (about 80% accuracy was discovered using CNN with some performance parameters). This comparison is being done to ascertain whether CNN is an accurate algorithm for this system in a real-time environment. A distance-measuring ultrasonic gadget that can also convert text into speech is used to provide information to those who are blind.

2.11. Real-time object recognition approach for assisting blind people (Zraqou et al., 2017)

The technique suggested by Zraqou et al. (2017) aids blind persons in recognizing their surroundings. By using tools like GPS services, they put it into practice by distinguishing closer and wider items that are attached to a blind person's glasses, ultrasonic sensors, and two cameras. These tools are used to gather the information they need around them. Researchers developed the dataset, and for feature extraction, they used the OpenCV library and SURF. For the purpose of sensing the items around them, they have utilized a real-time technique. The obtained accuracy is 90%. Therefore, being able to recognize objects with that level of accuracy can be advantageous. Future research aims to improve the ability of blind individuals to recognize more people in front of them while also explicitly knowing their names through face identification. It also aims to extract more information about objects, such as the color of a flower or the type of car in front of them.

2.12. Proposed platform

The fundamental issue with object detection is determining where an object is in a live broadcast and which category it belongs to. As a result, the workflow for an object detection model is largely split into three phases as per Figure 1.

Region selection: The entire image can be scanned using a multi-scale sliding window since distinct objects may appear across the image and have different aspect ratios or sizes. This thorough method has obvious shortcomings even if it can find every possible location for the pieces. Due to the vast number of candidate windows, it requires a lot of calculation and creates an excessive amount of duplicate windows. But if only a few sliding window templates are employed, unwanted zones might appear. While fast RCNN and faster RCNN can predict 200 and 300 region proposals for each image, respectively, YOLO can predict only about 100 region proposals per image.

Extraction of features: To detect various objects, we must take out visual components that can give a fair and accurate representation. A few of the distinguishing characteristics are SIFT (Jabnoun et al., 2015), HOG (Dalal & Triggs, 2005), and Haar-like (Lienhart & Maydt, 2002). This is due to the possibility that these traits could lead to representations associated with intricate brain cells. Manually creating an extensive feature descriptor is difficult that effectively characterizes all types of objects due to the huge variety of looks, lighting conditions, and backgrounds.

Figure 1
Object detection and identification system process and internal workings: a brief overview

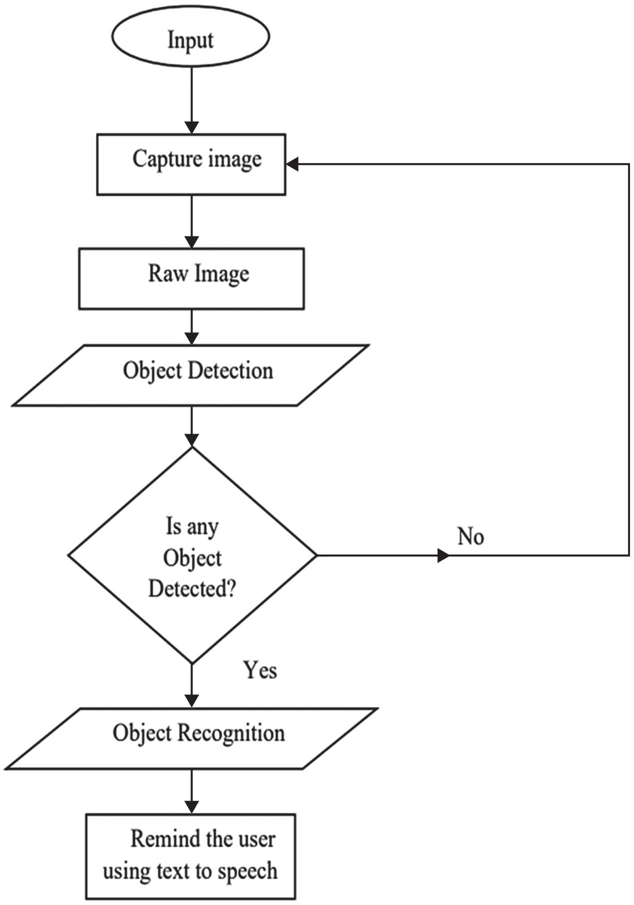
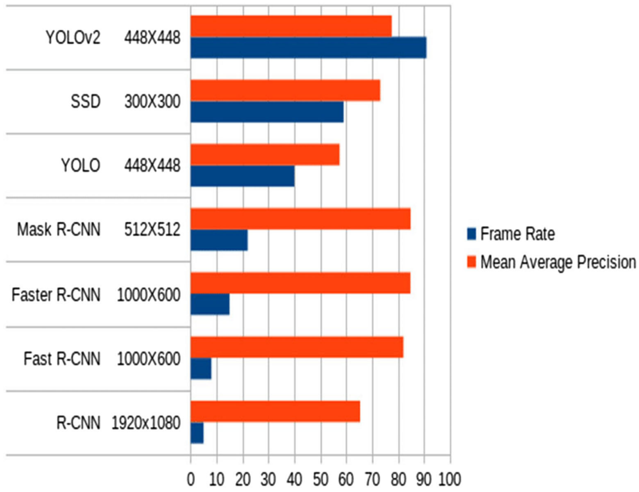


Figure 2
Performance comparison of various object detection models



Classification: In order to distinguish a target object from all other categories, a classifier is also necessary. Popular alternatives include support vector machine (Tzotsos & Argialas, 2008), AdaBoost, and deformable part-based model (DPM) (Felzenszwalb et al., 2009). The DPM is a flexible model among these classifiers that manages severe deformations by combining object components with deformation cost. In DPM, a graphical model is used to blend carefully defined low-level properties and kinematically inspired part decompositions. Furthermore, models based on high-precision parts can be created for different object classes via discriminative learning of graphical models.

3. Discussion

Object detection has gotten a lot of attention in recent times as one of the most basic and tough subjects in computer vision. After reviewing these many papers as per my view, as per Figure 2, there are a variety of algorithms for detecting objects, that is, region based and regression based, where region based is superior to regression based in terms of accuracy but speed is always given priority in real-time object detection systems. YOLO, SSD, RCNN, FAST RCNN, FASTER RCNN, etc. algorithms were used by researchers in their research in that YOLO is faster than fast RCNN or other region-based CNNs, but it is less accurate where the object is too close to the camera. Deep learning based object detection algorithms have been extensively used in a variety of diligence, but there are still several issues that need to be addressed that as Minimize overall calculation on data, to attain a high position of discovery effectiveness for small objects and multi-category object discovery is realized. For better results, DL needs a lot of data.

4. Conclusion

The effectiveness of the algorithm employed for detection is determined by its ability to find the object within an image. Using two benchmark datasets, including PASCAL VOC 2007 and Microsoft COCO, we examine several object detection techniques.

YOLOv3 was found to be the fastest, followed closely by SSD and Faster RCNN, which came in last. Even so, it is true that the use case affects the technique chosen; if you are working with a limited dataset and do not require real-time results, Faster RCNN is the best option. If you need to analyze a live video feed, choose YOLOv3. SSD, on the other hand, offers a good compromise between speed and accuracy. YOLOv3 is the most recent version of the three to be released, and it is actively being improved by the large open-source community. Hence, as a result of the out of the three object detection CNNs analyzed, YOLOv3 shows the best overall performance.

5. Future Scope

Object detection and recognition are most important for visually impaired people; researchers used many algorithms to work on it and make their life friendly but it has some drawbacks like less accuracy with smaller objects, and it may not perform well in cases where the object is too close to the camera so in future we can elaborate these drawbacks and enhance the performance and accuracy of algorithms.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Balaji, V., Raja, S. K. S., Raman, C. J., Priyadarshini, S., Priyanka, S., & Salai Kamalathai, S. P. (2020). Real time object detector for visually impaired using OpenCV. *European Journal of Molecular & Clinical Medicine*, *7*(4), 2963–2971.

Bastomi, R., Ariatama, F. P., Putri, L. Y. A. T., Saputra, S. W., Maulana, M. R., Syai'in, M., ... & Zuliari, E. A. (2019). Object detection and distance estimation tool for blind people using convolutional methods with stereovision. In *2019 International symposium on electronics and smart devices*, 1–5.

Bhole, S., & Dhok, A. (2020). Deep learning based object detection and recognition framework for the visually impaired. In *Fourth international conference on computing methodologies and communication*, 725–728.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer society conference on computer vision and pattern recognition,* 886–893.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139.

Jabnoun, H., Benzarti, F., & Amiri, H. (2015). Object detection and identification for blind people in video scene. In *2015 15th international conference on intelligent systems design and applications*, 363–367.

Joshi, N., Maurya, S., & Jain, S. (2021). Real-time object detection and identification for visually challenged people using mobile platforms. In *Workshop on advances in computational intelligence at ISIC,* 55–62.

Karmarkar, M. R. R., & Honmane, V. N. (2021). Object detection system for the blind with voice guidance. *International Journal of Engineering Applied Sciences 15 and Technology*, *6*(2), 67–70.

Lienhart, R., & Maydt, J. (2002). An extended set of Haar-like features for rapid object detection. In *International Conference on Image Processing*.

Lu, M., & Chen, L. (2020). Efficient object detection algorithm in kitchen appliance scene images based on deep learning. *Mathematical Problems in Engineering*, 1–12.

Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: State of the art. *Applied Intelligence*, *51*, 6400–6429.

Shah, S., Bandariya, J., Jain, G., Ghevariya, M., & Dastoor, S. (2019). CNN based auto assistance system as a boon for directing visually impaired person. In *2019 3rd International Conference on Trends in Electronics and Informatics*, 235–240.

Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data, 8*(1), 1–27.

Stephen, O., Mishra, D., & Sain, M. (2019). Real time object detection and multilingual speech synthesis. In *2019 10th International* conference on computing*, communication and networking technologies*, 1–3.

Suresh, A., Laha, D., Gaba, D., & Bhambri, S. (2019). Intelligent smart glass for visually impaired using deep learning machine vision techniques and robot operating system (ROS). In *Robot Intelligence Technology and Applications 5: Results from the 5th International Conference on Robot Intelligence Technology and Applications 5*, 99–112.

Tzotsos, A. & Argialas, D. P. (2008). Support vector machine classification for object- based image analysis. *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 663–677.

Vaidya, S., Shah, N., Shah, N., & Shankarmani, R. (2020). Real-time object detection for visually challenged people. In *2020 4th International conference on intelligent computing and control systems*, 311–316.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, *30*(11), 3212–3232.

Zraqou, J. S., Alkhadour, W. M., & Siam, M. Z. (2017). Real-time objects recognition approach for assisting blind people. *International Journal of Current Engineering and Technology*, *7*(1), 2347–5161.

How to Cite: Katkade, S. N., Bagal, V. C., Manza, R. R., & Yannawar, P. L. (2023). Advances in Real-Time Object Detection and Information Retrieval: A Review. *Artificial Intelligence and Applications* 1(3), 139–144, https://doi.org/10.47852/bonviewAIA3202456