International Journal of Current Research and Review
DOI: http://dx.doi.org/10.31782/IJCRR.2020.122032

Scopus

Research Article

IJCRR
Section: Healthcare
Sci. Journal Impact Factor: 6.1 (2018)
ICV: 90.90 (2018)

Object Detection Using Machine Learning for Visually Impaired People

Venkata Naresh Mandhala[1], Debnath Bhattacharyya[1], Vamsi B.[2], Thirupathi Rao N.[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India; [2]Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Vishakhapatnam, AP, India.

ABSTRACT

In this challenging evolution, the primary task in detecting the objects requires a computer vision that deals over indoor and outdoor classes. Over the past decades, this zeal requires more attentiveness. Previous implementation techniques involve in object detection with a strategy of single labelling.

Aim and Objectives: In this regard, a multi-label approach using machine learning and vision technologies, and accurate response can be acknowledged based on its accuracy and effectiveness. In the proposed work, we solve the existing system problem by using classification/clustering techniques that are used to reduce the recognize time of multi objects in less time with best time complexities.

Model: The model used to assist the visually impaired people can independently recognize objects which are near to them. The reverence, combined with the study, confounded the inception of these machine learning algorithms for visually impaired persons in assisting the accurate navigation, including indoor and outdoor circumstances.

Conclusion: In this connection, an indoor and outdoor architecture on Retina Net is implemented for its detection techniques, and also neural network technologies support this framework. Based on the effectiveness and implementation time, ResNet and FPN act as a crucial module for its accuracy.

Key Words: Object Detection, Machine Learning, RetinaNet, Yolo, Visually Impaired People

INTRODUCTION

Computer Vision

For analyzing the Visual world to break and elucidate, which explains computer vision in computer technology. In categorizing the objects' accuracy, machines use deep learning models[17] and digital images such as cameras and videos. In the early 1950s, demonstrations have already started in computer vision to identify the keen edges and align the simpler objects with falling under categories such as circles and squares by the techniques of first neural networks. Later in the 1970s, Optional character recognition came into existence of computer vision explicated typed or handwritten data on its primary trading tool. The illustrated data mainly used for the blind as a development.[3] In the 1990s, the World Wide Web has evolved, producing sizeable images for examining and various computing facial detection had developed. These evolving text frames supported the analysis of machines in detecting particular persons in pictures and videos.[20,21,24]

The image segmentation has to be inspected individually by categorized into various partitions or frames. The object detection indicates detecting a particular object in the image. Upgraded object detection admits multiple objects in a single image. For example, in certain instances, like the football field, an offensive player, a defensive player, a ball, etc. To obtain this X, the Y Coordinate model is implemented for the bounding box and detecting everything inside the region.[6,7] The facial recognition technology for Object detection has come up with the latest type that concedes the human face in the entire image and detects as a person in particular. Using pattern detection, a duplication of the shapes, colours, and other visual indicators in the picture. The image classifications are used to bring together into multiple divisions. The

Corresponding Author:

Venkata Naresh Mandhala, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India; Email: mvnaresh.mca@gmail.com

ISSN: 2231-2196 (Print) **ISSN:** 0975-5241 (Online)

Received: 08.07.2020 **Revised:** 15.08.2020 **Accepted:** 25.09.2020 **Published:** 27.10.2020

feature similarities can be attached to pattern detection that classifies the similitude among the matched objects.[19]

The Staggering evolution of the computer vision includes these advanced effects. Around 50 to 90 %, the accuracy in detecting the objects and dividing them into specific categories has rapidly been increased in less than a decade. In the day-to-day usage of computers, they maintain keen exactness in detecting and reacting to visual inputs compared to humans.

The computer vision can be differentiated in three stages for real automation.

a) **Stage 1:** On examination 3D technology, which obtains a synchronous view over images and even large sets.[2]
b) **Stage 2:** Thousands of labelled or pre-identified images are automated models trained by deep learning models.
c) **Stage 3:** The ending step is the elucidative step in which the object is detected or arranged.

Machine Learning

Analytical model building is computerized by the method of data examination in Machine learning. It deals with the department related to artificial intelligence hinge on the machine, can understand the data, detect design, and also make resolutions with even human involvement.[23] These machine learning charges are broadly divided into numerous divisions, as shown in **Figure 1**, namely supervised learning and unsupervised learning.

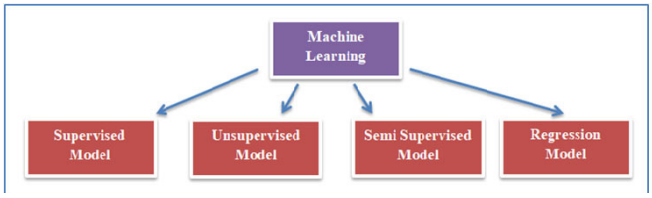

Figure 1: Machine Learning Models.

Supervised Model: These algorithms were instructed using labelled instances, such as retrieving the required output by giving specified inputs. An example, such as detecting the online marketing of products, the outcome would be either "order delivered successfully" or "not delivered," in which it is mentioned in the system language of Boolean values either true or false. This learning algorithm retrieves accurate outputs proportional to the group of inputs they receive. This algorithm also identifies the exact outputs to the original outputs on collations for the detection of the errors, and it alters the algorithm consistently. There are also supervised learning techniques that use specific designs to forecast the values of the label on additional unlabeled data, namely classification, regression, prediction, and gradient boosting. In an application such as foretelling historical data like upcoming events, this supervised learning is frequently used. For instance, it precedes in deceitful cases like credit card transactions or when an insurance customer is at a point in filing a claim. The Regression algorithms include any value within a specific collection that produces uninterrupted outputs. For example, uninterrupted outputs are the values like temperature, length, or price of the object .[26]

Unsupervised Model: These algorithms develop an algebraic model form the group of data, which takes only inputs irrespective of the output flags. The assembling or congregation of data tips to attain a beautiful structure in the data obtained through this unsupervised learning algorithm. As in attribute learning, inputs grouped into divisions by collecting patterns in the data can also be achieved through unsupervised learning. The procedure involving reducing the count of "features," or the inputs among the data set, is often termed to as dimensionality reduction.

Another learning mechanism that uses identical implementations as supervised learning is referred to as Semi-supervised learning. Priming can be handled in labelled and unlabelled data – Expendently smaller amounts of labelled data with many unlabelled data (Unlabelled data is cheaper and essential for lesser endeavours to gain). Classification, regression, and prediction methods are implemented in this part of learning. Semi-supervised learning is applicable or maintained to its fully labelled training process only when the cost is related to it too high. An example of this type of knowledge is detecting an individual's face in front of a web camera.

The significant role in monitoring the patient's health in health care instantaneously by providing wearable devices and sensors in this vast developing health industry. It is also used in diagnosis and treatment methods to detect the courses or errors in the medical expert's analysis. In government sectors, vast data is accessed through different sources that can be detected for the perception that needs to be prevented and made safer by the tools that mainly need machine learning technology. Discovering the confidential data, for instance, finds ways to increase its accuracy and save money. Identifying the deception and reducing detection frauds can also be maintained through machine learning.

The critical role of machine learning is to understand better the data's configuration similar to the statistical models – to perceive better about data theoretical distributions are used. The mathematical model theory is proved mathematically based on its model; provided that the machine learning should clarify the assumptions the data should meet. This has been instrumented on the data that holds accurate solutions while using computers to verify the configuration in which the data is maintained. No theory had explained the formation of structure in its appearance. The machine learning undergoes testing in which it only relies

upon affirmation of errors on the new data, but doesn't focus on the theoretical test which returns null proportions. This is because learning can be made merely automatically but often use a repetitive approach in understanding the data. These will iterate through detailed data until a required structure is obtained.

Neural Networks

Neural Networks are closely related to the brain's structure since it uses computing systems with interconnected nodes like the neurons in the human brain. For obtaining the concealed designs and associations among the raw data, algorithms are used to cluster and classify them and produce constant learning and improvements.

This neural network's primary goal is to develop a ciphering system that functions the same as that of the human brain in technically solving the problems. Furthermore, it has been handled under the classification of colliding some distinct tasks at certain times, which are mainly used in botanical applications in particular divisions. From then, various works combining the vision of the computer, recognition of speech, machine translation, social network clarifications, playing boards, and video games and medical diagnoses are maintained by using these neural networks.

Types of neural networks: The different kinds of deep neural networks are shown in figure 2.

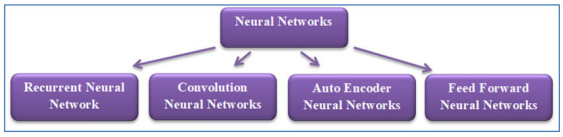

Figure 2 : Neural Network Categories

Recurrent neural networks (RNNs): This following information, namely, time-stamped data from a sensor device or a sentence that is spoken, is determined in the terms sequentially. In contrast to traditional neural networks, the inputs to the recurrent neural networks do not depend on one another; instead depends upon the calculations of its other elements to give the desired outputs. Forecasting and time series applications, sentiment analysis, and other text applications are applications that involve RNNs.

Convolution neural networks (CNNs)*:* This includes five types of Layers, namely: input, convolution, pooling, fully connected, and output. Every layer had its specialization, like summarizing, connecting, or activating. These CNNs can also be used in various other studies, such as natural language processing and forecasting.

Autoencoder neural networks: These are formed through some preoccupations, namely encoders, formed by a group of inputs given. Conventionally, autoencoders mainly use an unsupervised model in which it is evaluated on itself, comparing it with the other traditional neural networks. The encoders are used to render insensible data to irrelevant and sensible data the relevant based on the model. As this is a layered model, further generalization can accumulate from the higher layers (when a decoder enables to the nearest point). These assumptions can be of linear or nonlinear extensions or categories.

Feedforward neural networks: This network connects the layer one over the other in a sequential way along with the other layer's connections. The data in this network is unidirectional in which the data can be taken from one to another in progressive direction only. There will not be any loops related to feedbacks.

Object Detection

Based on the lightening conditions, some of the objects, particularly in the indoor case, the image might not be visualized depending upon the pixels that have been detected by the system.[12] In lightening conditions, constituting the captured image input, which is not recognized by the system, has become the most challenging task. Putting these issues under consideration, the work has become very challenging and dense. The potential of traversing from one location to another is that which is related to our day to day lives.

For a computer to understand circumstances to identify and discover the objects in images and videos plays a significant task. As for humans, this is very simple when compared to that of the system to be recognized. In the visually impaired persons, this has to become a challenging task in solving the discrimination.

The object detection algorithm needs to take out the features pertained to certain specific classes; there might be a large quantity of pre- interpretation is required. Detecting and recognizing the object particular to the indoor captured image, which is taken as an input image by the system, should also detect the adjoining environments in its surroundings.

Literature Survey

Mouna and Riadh have proposed their work on "An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation" in 2020. In this work, the function of computer vision is to detect indoor objects accurately. The visually impaired people can be assisted by navigating the purposes of the CNN framework.[4,5,14] To identify the specific objects first, we need to detect the pixels available in the images. If the lighting conditions are wrong, then it is challenging to capture and identify the objects with high accuracy. To detect the indoor objects, the algorithm needs to extract the image features with a particular class, and it can be done by RetinaNet.[25] To enable the network for small object detection by a Region Proposal

Networks (RPN), which involves subsampling to obtain the image information. The Resort with 152 samples achieved an average precision with 83.1%, and DenseNet with 121 samples achieved an average precision with 79.8%.

Han Hu and Jiayuan have proposed their work on "Relation Networks for Object Detection" in 2018. Based on the relation models, this work assigned an equal quantity of work by considering its features. This removes duplication and attains accuracy at specific standards. Since the objects are aligned in the 2D scale ratio, it uses objects rather than words. Further, the model is categorized into two components that fall under geometric and original weights.[15]

Xiangrong and Alan have proposed their work on "A Time-Efficient Cascade for Real-Time Object Detection: With applications for the visually impaired" in 2005. In this work, the main objective is to focus mainly on time complexities and their accuracies depending upon the various test that has been performed by the greedy approach the module which detects the text in the images which can be improved for visually impaired people. The quality of the model can be measured by F.P. and F.N. rates. The decision capability of the algorithm can be done by a set of training images and classifiers. The smart telescopic system will be used for vision problem people. On the micro screen visuals, the image represents itself in a emphasize way leaving certain spots of the image behind.[1]

Alice Tang and Zhiyuan have proposed their work on "Automatic Registration of Serial Cerebral Angiography: A Comparative Review" in 2018. During times, based on this work mainly in the medical field, specific changes have been made in identifying the disease and recoveries by considering its effectiveness and accuracy. Magnetic resonance imaging (MRI) and computed tomography (C.T). are analyzed on image processing algorithms that are highly examined rather than DSA. While DSA is referred to the diagnosis of several neurovascular conditions which is used at the time of surgeries, on these considerations, it could be concluded that the framework is designed based upon the patients diagnosed with ischemic stroke.[18]

Wei and Xia have proposed their work on "HCP: A Flexible Convolutional Neural Network (CNN) Framework for Multi-label Image Classification" in 2015. In this work, a CNN model produces the best performance for image classification with a single label. Due to complexity, multi labelling is an open challenge for training image layouts. A single image object is taken as an input will be given for hypotheses extraction, and this is shared with CNN to get individual scores by max pooling. The image's hypotheses are identified with different colours that can be indicated by different clusters.[10] The extraction method produces predictive results that are utilized by max pooling. By comparing the I-FT and HCP models, the HCP model improves the system performance by 5.7%.

Rim and Issam have proposed their work on "Indoor Object Recognition in a combination of a RGB(Red-Green-Blue) image and its corresponding depth image (RGBD) Images with Complex-ValuedNeural Networks for Visually-Impaired People" in 2018. In this work, the multi-model is used for visually impaired people to detect the objects with a multi-class strategy in an indoor area. This model takes at a time more than one label. The CVNN and multi-label techniques associate the image with labels that correspond to categories of objects at once.[16] The clusters can be made based on multi labelling by ML-CVNN, and the L-CVNN method works by image transformation to classify the problem by ranking solution. The input strategy captures the image by multi-label and multi classes to generate the contexts of realistic and non-realistic of nested and exclusive structures.[11,22]

Liang and Miachel have proposed their work on "Using multi-label classification for acoustic pattern detection and assisting bird species surveys" in 2016. This model is used for detecting the patterns in urban areas such as public streets, raining, restaurants, etc.[13] This method characterizes the audio clips, which yields the patterns. The main limitation of this model is to require a trained data set.

In 2012, the proposal of identifying the objects in today's day with data sets containing 1 million frames using a camera had already been initiated by Pirsiavash and D. Ramanan on "Detecting activities of daily living in First-person Camera Views." It can be done by the mapping of a set of analyzed frames with objects. Each object can found the maximum value by dividing the raw scores. This model does a cross-validation process for both tests and trained data to detect and remove the duplicates objects. To achieve accuracy for object detection confusion matrix is used for evaluating the classifier errors. The prepared data set contains 24 categories of objects with 1200 labels.[9]

Mohamed and Farid have proposed their work on "A Compressive Sensing Approach to Describe Indoor Scenes for Blind People" in 2015. In this work, the objects can be detected by bind people through a camera in different indoor environments. This model works by multi labelling strategy to measure the Euclidean distance and Gaussian Process. It verifies the presence of various objects concerning the data set. If the presence of particular objects is found, then it also sees the positions.[12]

Yong Lee and Ghosh have proposed their work on "Discovering important people and objects for egocentric video summarization" in 2012. In this work, the Camera wearer's day is a dense storyboard briefing the recommended methods. On the other hand, in traditional essential chunk selection techniques, the final presentation of these techniques mainly examines the vital objects and people who interact using this camera wearer. A few chucks/data packets required for the storyboard are reflected by the vital object-driven circum-

stances in this method. Based on our practices 17 hours of self-centred data depending upon the existing techniques, it shows excellence in saliency and summarisation. This has been done in 4 main steps; they are: (a) the image about a famous person or object could be predicted using a novel self-centric saliency cues which it trains a group independent regression model. (b) Separation of each task/event in dividing the video into subcategories of tasks. (c) Gaining the importance of each event by enabling the regression mechanism. (d)Choosing respective critical data chunks for the storyboard depending upon the required people and objects for representations.[8]

System Architecture

The architecture has been bought out, keeping in view that general algorithms are usually implemented in OpenCV, referred to as the popular computer vision library. Past evolutions on these theories of object identification also used these algorithms. These evolving technologies in the building of the latest applications based on these algorithms are not so useful and accurate. Hence, these traditional algorithms could not meet its specifications in evaluating its performance and work efficiency under certain circumstances.

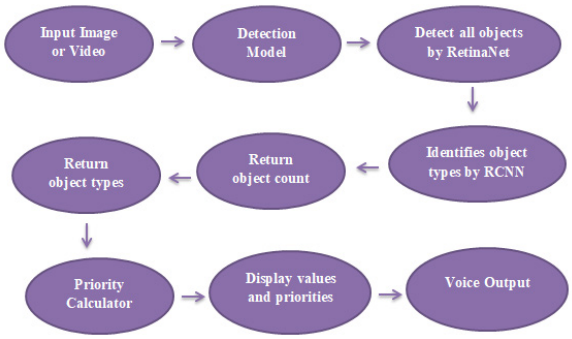

Figure 3: Architecture to detect the objects.

As shown in **Figure 3,** the object detection model, and the video object detection model captures image dependencies. Image TK can do the user interface dependencies, and the gTTS package can do Tkinter packages and the voice output dependencies. The detection model loads the image from the specified locations. The image grid labelling cab is set for column=1 and row=2. The loaded image seized with dimensions of 665 and 490. The support in verifying the video type formats can be examined on a graph, adjusting the XY plane to x=1 and y=1 in a multi labelling model. The detected image is given to the RetinaNet algorithm to train the model to detect all the objects available. The prepared data is provided to RCNN for identifying all the types of objects. The probability percentage for each purpose is calculated with a precision value of more than 15. The priority calculator is estimating the accuracies of each object based on the count. The voice output is an efficient idea work done by this model useful for visually impaired people. It contains a message to say the numbers of objects are available in front of the person. This model detects two frames per second, and the minimum percentage probability is 30.

Tensor Flow:

The Computer Vision Python library uses a simple ImageAI that encourages developers to combine state-of-the-art Artificial Intelligence features to its subsist and provisional applications and systems.

In Machine language, the tender flow acts as an end-to-end open-source platform. The comprehensive, flexible ecosystem of tools, libraries, and community resources allows the developers for the state-of-the-art in ML and could quickly build and evaluate ML-powered applications.

Open CV:

The open CV included in the machine learning software library is also an open-source computer vision. In commercial products, it is used as a general infrastructure for computer vision applications and speeds up the use of machine perception.

Object detection model:

Figure 4 shows that the object detection model initiates the workflow of the model. This model depends on input as an image and output as a voice message. First, the grid is set with values column=1 and row=2. After setting the grid size, configure the image for displaying the path. The image can be re seized with 665 X 490 dimensions for setting at the user interface. Labelling the image with parameters x=1 and y=1, after loading an image completed. The 'object detection' model initially verifies the type of image file for loading the model, input, and output paths.

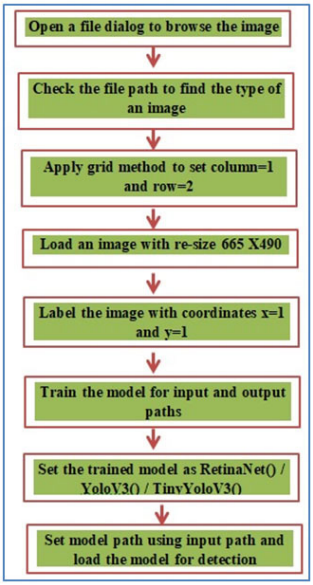

Figure 4: Object detection model.

Algorithm: ObjectDetection

Step 1: Load the trained model

Step 2: if (setModel=RetinaNet( ) || YoloV3( )|| TinyYoloV3( )){

Step 3: for eachObject in setModel do {

Step 4: {

Step 5: if(probability(eachObject)>15)

Step 6: {

Step 7: update(eachObjectName);

Step 8: update(percentageProbability);

Step 9: }

Step 10: detectedImage();

Step 11: }

Algorithm: Detected Image

Step 1: Open the input image path

Step 2: Resize the image with 665 X 540 dimensions

Step 3: Label the image with parameters x:=2 and y:=55

Step 4: for each box in a frame do

{

x:=5, y:=5, width:=675, height:=100

}

Step 5: for each frame in detection do

{

x:=5, y:=615, width:=675, height:=85

}

Step 6: for each frame in the detected image do

{

x:=685, y:=5, width:=680, height:=605

}

Step 7: for each frame in voice do

{

x:=685, y:=615, width:=680, height:=85

}

Yolo Algorithm:

Object detection can be identified using convolution neural networks as an algorithm. Detection algorithms can be compared as identification of the image can be done not only by identifying the label of a particular class but also identifies its location in which the object is placed. This algorithm also helps to divide the picture into parts but also in identifying different objects that are located along with the image. This convolution neural network algorithm uses only a single structure neural network to detect the complete image by dividing the image into separate portions and identifying closed boxes and probabilities for separate portions. These bounding boxes in the image are calculated using pre-identified probabilities.

The image which has been kept under identifying the objects uses YOLO as the substitute for its size. In general, we consider only fixed size due to different issues that only reveal its main objectives while evaluating the detection process by using algorithms.

This algorithm initially fixes the image's height and width as input and gets the output. It lists out all the boxes available in the frame with a class multi labelling. Each box in the frame contains *(p_p,bx, by,bh,bw,p)* as a parameter where 'p_p' can be either 0 or 1 which defines the probability of a person '*p*' is present in the image or not, '*bx*' and '*by*' defines the midpoint of the box and '*bh*,' '*by*' defines the height and width of the box respectively.

Working procedure of Yolo V3:

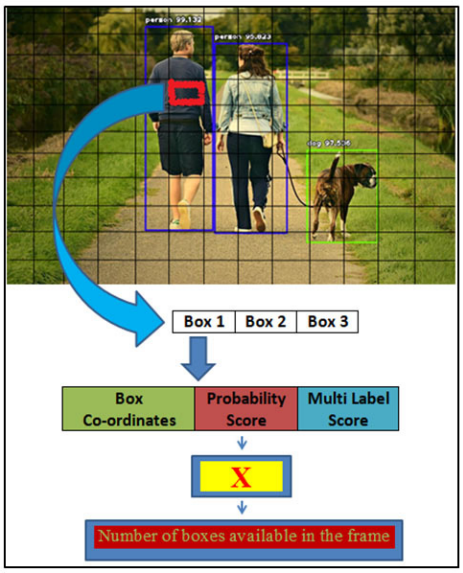

Figure 5: Workflow of Yolo V3.

Figure 5 shows the workflow procedure of Yolo V3. In this, the feature prediction mapping is done on each box available in the frame. Each box is treated three individual boxes, which defines V3. Each box's attributes contain 'box coordinates,' 'probability scores,' and 'multi-label scores' for bounding all the boxes available in the frame. The 'red'

colour cell is at the 5th row of the sixth cell on the grid image. Now we have applied the feature mapping on it to detect the person.

Algorithm:Yolo

Step 1: if (setModel=YoloV3( )|| TinyYoloV3( ))

Step 2: Set the execution path

Step 3: Set the model path to load the model

Step 4: Set timer as default timer to get the elapsed time

Step 5: Detect the objects from input to output image

Step 6: join the detector on inputImage and outputImage

Step 7: obj:=0, accur:=0

Step 8: for each object in detection do

Step 9: {

Step 10 :obj:=obj+1; accr:=accr+percentprob;

Step 11: }

Step 12: eTime: = defaultTimer – startTime;

After applying the frame's echoing technique to get the two-dimensional shape, we have computed the probability score for each box by applying the product. The sample example on how to get the probability scores as follows:

$$P_{score} = P_P * \begin{Bmatrix} Classlabel1 \\ Classlabel2 \\ Classlabel3 \\ Classlabel4 \\ . \\ . \\ . \\ Classlabeln \end{Bmatrix} = \begin{Bmatrix} Pp * Classlabel1 \\ Pp * Classlabel2 \\ Pp * Classlabel3 \\ Pp * Classlabel4 \\ . \\ . \\ . \\ Pp * Classlabeln \end{Bmatrix} = \max \begin{Bmatrix} 91.8 \\ 99.1 \\ 94.8 \\ 98.2 \\ . \\ . \\ . \\ 90.3 \end{Bmatrix} = 99.1$$

RetinaNet Algorithm

Based on the Computer Vision Python library, the Retina Net acts as unique object identification. A group of networks that serves as a significant backbone of such a system will act as two task-specific sub-networks. The Retina Net has a one-stage detector that uses focal loss/lower loss, relying on "simple" negative testing samples. On the other hand, the loss is focused on "complex" examples, too, enhancing the identification accurately. ResNet+FPN (Feature Pyramid Network) acts as a significant part for feature descent, plus two task-specific sub-networks for dividing into categories and bounding box regression, which combines to form the Retina Net. Enhancement of the Retina Net is obtained from the two progressive changes over the existing single-stage object detection models.

Algorithm: Retina Net

Step 1: if (setModel=RetinaNet( ))

Step 2: Set the execution path

Step 3: Set the model path to load the model

Step 4: Set timer as the default timer to get the elapsed time

Step 5: Detect the objects from input to output image

Step 6: join the detector on the input image and output image

Step 7: obj:=0, occur:=0

Step 8: for each object in detection do

Step 9: {

Step 10 :obj:=obj+1; accr :=accr+percentprob;

Step 11: }

Step 12: eTime: = defaultTimer – startTime;

Feature Pyramid Network (FPN) acts as building a rich multiscale feature pyramid from one single resolution input image, located on the upper layer of the ResNet. FPN is multiscale, semantically strong at all scales, and fast to compute. With the help of Focal Loss (FL), the training set provides a high improvement performance by limiting the relative loss.

$$FL(\rho_t) = -\propto_t (1-\rho)^{\gamma} \log(\rho_t)$$

Pyramid networks almost use standard in identifying the objects at distinct levels. On the other hand, the inherent multiscale pyramidal hierarchy of deep CNNs in creating feature pyramids Feature Pyramid Network (FPN) is used. Figure 6 clearly shows ResNet's usage, pyramid networks, class subnet, and box subnet methods in stage-wise architecture.

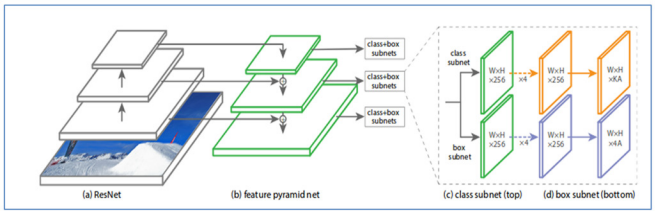

Figure 6: Stage of smart Architecture.

RESULTS AND DISCUSSION

The image or video can be loaded into the object detection model, as shown in figure 7. This interface contains loading an image, running a module to execute the program, the number of detected images detected in the module and play audio for better understanding for visually impaired persons.

Figure 7: Interface to run the detection modules.

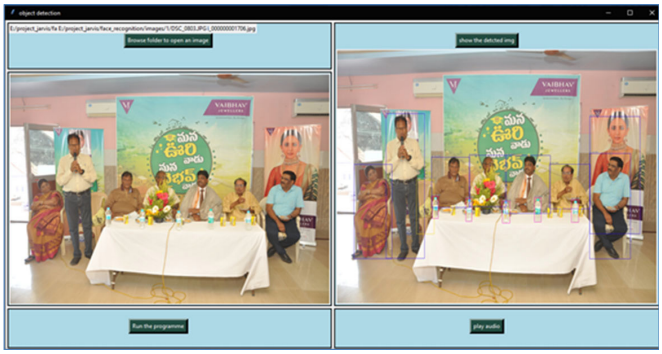

Figure 8: Object detection in an outdoor environment with multi labelling.

Figure 8 shows the loaded image in the outdoor environment at one part, and the other hands, the model is marked with all the objects available in the picture with blue-coloured frames.

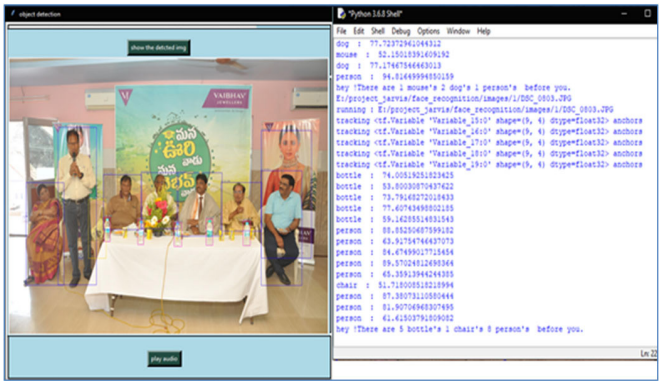

Figure 9: Accuracy values of available objects in the image.

Figure 9 shows all the available objects in the images with accuracies. The detection module observed that there are five bottles, one chair, and eight persons are detected in the loaded image. By playing the audio, the module says, "Hey! There are five bottles, one chair's eight people's before you".

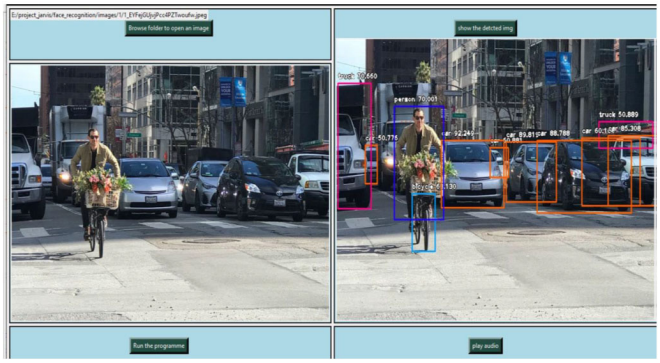

Figure 10: Object detection in a traffic environment.

Figure 10 shows all the available variables with labels at the traffic signal environment. In this case, the model detected seven cars, two trucks, one person's, and one bicycle's before the person, along with accuracy.

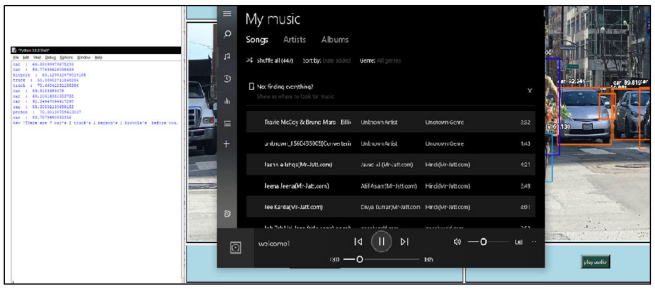

Figure 11: Playing audio output.

Figure 11 shows the accuracy of the objects available in the loaded image. By playing the "play audio" module, the visually impaired people can listen to the type of objects in the surrounding environment and the count.

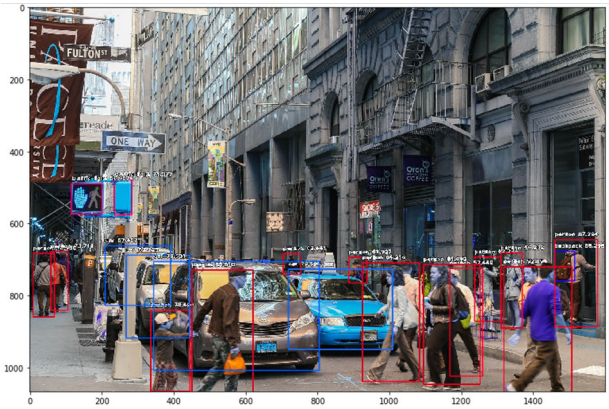

Figure 12: Object detection at a traffic signal.

Figure 12 shows all the detected variables applied at a traffic signal and used for the comparative analysis to compare results by Retina Net, Yolo V3, and Yolo Tiny on the same image.

Table 1: Overall objects detection done by model on different images

Algorithm	Number of Frames	Identified objects	Accuracy
Retina Net	4	Dog-1	77.7
		Mouse	52.1
		Dog-2	77.1
		Person	94.8
Yolo V3	14	Bottle-1	74.0
		Bottle-2	53.8
		Bottle-3	73.7
		Bottle-4	77.6
		Bottle-5	59.1
		Person-1	88.8
		Person-2	63.9
		Person-3	84.6
		Person-4	89.5
		Person-5	65.3
		Person-6	87.3
		Person-7	81.9
		Person-8	61.6
Yolo Tiny	3	Chair	51.7
		Sofa	87.5
		Chair-1	75.3
		Person	92.6
		Chair-2	78.3

Table 1 shows the results applied to different images given by Retina Net, Yolo V3, and Yolo Tiny. Based on the trained data set, the Retina Net has the highest accuracy of 94.8% for detecting the person, but accuracy for detecting the animals is moderate because these objects are available at the corners. For detecting the Yolo algorithm's person, the 'Yolo Tiny' has given the highest accuracy compared to 'Yolo V3'.

Table 2: Comparision over indoor and outdoor detection objects on different images

Type	Algorithm	Number of frames	Object	Accuracy
Indoor	Retina Net	3	Person-X	98.4
			Person-Y	95.6
			Dog	77.7
Indoor	Yolo Tiny	4	Sofa	86.3
			Chair-X	75.3
			Person	92.6
			Chair-Y	78.3
Outdoor	Retina Net	6	Car-X	60.8
			Car-Y	50.7
			Truck-X	50.8
			Truck-Y	70.6
			Car-Z	88.7
			Person	70.0
Outdoor	Yolo Tiny	4	Bottle-X	77.6
			Bottle-Y	59.1
			Person-X	88.8
			Person-Y	63.9

Table 2 shows the results applied to different images in different environments, such as indoor and outdoor. The 'Retina Net' has the highest accuracy in the indoor environment compared to the 'Yolo Tiny.' In an outdoor environment, both 'Retina Net' and 'Yolo Tiny' have intermediate results. To get a better clarification on the algorithm model, we measured the accuracy based on the measurement parameter, as shown in table 3.

Table 3: Comparision over a distance of objects on different images

Type	Algorithm	Object	Measurement	Accuracy
Outdoor	Retina Net	Car	Milli meters	92.6
	Retina Net		Meters	60.1
	Yolo V3		Meters	50.7
Outdoor	Yolo Tiny	Truck	Meters	50.8
	Retina Net		Meters	70.6
Outdoor	Yolo V3	Person	Meters	70
	Yolo Tiny	Car	Meters	80.7
Indoor	Retina Net	Person-X	Milli meters	87.3
	Retina Net	Person-Y	Milli meters	81.5
	Yolo V3	Chair	Meters	51.7
Indoor	Retina Net	Bottle-X	Meters	73.7
	Yolo V3	Table	Meters	51.9
	Yolo Tiny	Bottle-Y	Meters	77.6
	Yolo Tiny	Bottle-Z	Meters	78.6
Indoor	Retina Net	Person-X	Milli meters	94.8
	Retina Net	Person-Y	Milli meters	98.6
	Yolo Tiny	Dog	Milli meters	96.4

Table 3 shows the results applied to different images over distance measurement at indoor and outdoor. In both environments, the 'Retina Net' algorithm has the highest accuracy over 'Yolo Tiny' and 'Yolo V3' over the distance measurement in 'meters' and 'millimetres.'

Table 4: Comparision of Retina Net, Yolo Tiny and Yolo V3 on the same image

Objects	Retina Net Algorithm	Yolo V3 Algorithm	Yolo Tiny Algorithm
Backpack	56.8	68.2	0
Traffic Light	59.1	71.3	66.7
Traffic Light	80.8	56.9	0
Car-1	82.8	57.4	61.9
Car-2	81.2	62.2	84.5
Car-3	82.8	79.5	51.2
Car-4	89.5	79.5	61.5
Person-1	53.9	62.4	64.1
Person-2	66.6	61.9	75.2
Person-3	65.6	70.9	81.7
Person-4	73.5	73.7	83.7
Person-5	74.1	62.2	91.9
Person-6	77.2	56.2	94.3
Person-7	78.1	87.2	98.4
Person-8	84.3	52.4	0
Person-9	92.3	78.4	0
Person-10	96.6	83.6	0
Person-11	97.9	84.2	0
Person-12	98.5	81.1	0
Person-13	99.1	92.9	0
Person-14	99.9	0	0
Objects Detected	**21**	**20**	**12**
Average Accuracy	**80.5**	**71.1**	**76.3**
Elapsed Time in sec	**6.62**	**9.40**	6.25

Table 4 shows the compared results among all three algorithms, such as Retina Net, Yolo V3, and Yolo Tiny. After applying these algorithms in figure 10, it is observed that the 'Retina Net' algorithm has detected most objects which are available in the image. By comparing 'Yolo V3' and 'Yolo Tiny,' the 'Yolo Tiny' has given more accurate results than 'Yolo V3, but it is detected only 12 objects and in case of 'Yolo V3' it has detected 20 objects with an accuracy of 71.1%. So in real-time, detecting fewer objects is not acceptable for visually impaired people. The 'Retina Net' algorithm has given the highest and best accurate results among the remaining algorithm with less elapsed time.

The graphical representation of compared results available in Table 4, as shown in figure 11, with respective of the number of objects detected and the average accuracy of all objects.

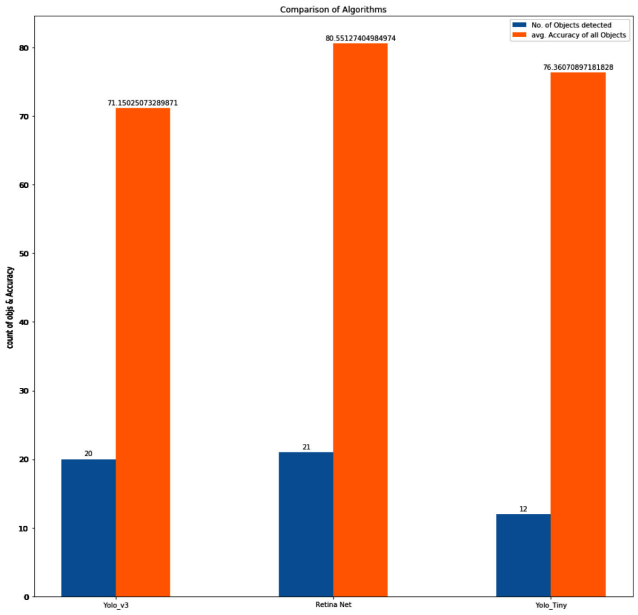

Figure 13: Accuracy Comparision of algorithms.

CONCLUSION

Although Objection detection is an esteemed task yet, it is an innovative errand. It plays an essential role in numerous implementations like identifying an image, auto-annotation of image, and apprehension of the ideology. Eliminating the problem of vision in visually impaired persons, the proposed work can be used effectively in detecting the objects along with their design patterns in an exact manner and to identify them among multiple different objects in a captured input image individually with high accuracy and with expert navigation, by implementing the Specific model X-Y plane by calculating their percentages accurately of the detection and also supporting the transformation input images to speech. The object detection also furnishes its results on multiple objects and various methodologies in discovering artefacts, identifying and collating each step for its productiveness.

ACKNOWLEDGMENT

Authors acknowledge the immense help received from the scholars whose articles are cited and included in references to this manuscript. The authors are also grateful to authors / editors / publishers of all those articles, journals, and books from which the literature for this article has been reviewed and discussed.

Conflict of Interest: Nil

Source of Funding: Nil

REFERENCES

1. Chen X, Yuille AL. A time-efficient cascade for real-time object detection: With applications for the visually impaired. In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops 2005 Sep 21:28-28.
2. Tatsuro UE, Hirohiko K, Tetsuo T, Akihisa O, Shin'ich Y. Visual Information Assist System Using 3D SOKUIKI Sensor for Blind People. the 32nd annual of the IEEE Industrial Electronics Society (IECON). 2006.
3. Hub A, Hartter T, Ertl T. Interactive tracking of movable objects for the blind on the basis of environment models and perception-oriented object recognition methods. InProceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility 2006 Oct 23:111-118.
4. Ando B. A smart multisensor approach to assist blind people in specific urban navigation tasks. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2008 Aug 15;16(6):592-4.
5. Andò B, Graziani S. Multisensor strategies to assist blind people: A clear-path indicator. IEEE Transactions on Instrumentation and Measurement. 2009 Apr 24;58(8):2488-94.
6. Yang X, Tian Y. Robust door detection in unfamiliar environments by combining edge and corner features. In2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops 2010 Jun 13:57-64
7. Hasanuzzaman FM, Yang X, Tian Y. Robust and effective component-based banknote recognition for the blind. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2012 Jan 18;42(6):1021-30.
8. Lee YJ, Ghosh J, Grauman K. Discovering important people and objects for egocentric video summarization. In2012 IEEE conference on computer vision and pattern recognition 2012 Jun 16: 1346-1353.
9. Pirsiavash H, Ramanan D. Detecting activities of daily living in first-person camera views. In2012 IEEE conference on computer vision and pattern recognition 2012 Jun 16:2847-2854.
10. Wei Y, Xia W, Lin M, Huang J, Ni B, Dong J, Zhao Y, Yan S. HCP: A flexible CNN framework for multi-label image classification. IEEE transactions on pattern analysis and machine intelligence. 2015 Oct 26;38(9):1901-7.
11. Cadena C, Dick A, Reid ID. A fast, modular scene understanding system using context-aware object detection. In2015 IEEE International Conference on Robotics and Automation (ICRA) 2015 May 26:4859-4866.
12. Mekhalfi ML, Melgani F, Bazi Y, Alajlan N. A compressive sensing approach to describe indoor scenes for blind people. IEEE Transactions on Circuits and Systems for Video Technology. 2014 Nov 20;25(7):1246-57.
13. Zhang L, Towsey M, Xie J, Zhang J, Roe P. Using multi-label classification for acoustic pattern detection and assisting bird species surveys. Applied Acoustics. 2016 Sep 1;110:91-8.
14. Vasireddy S, Ravipati V, Ravi T, Jegan G. Wireless sensor-based GPS mobile application for blind people navigation. ARPN Journal of Engineering and Applied Sciences. 2016;11(13):8374-9.
15. Hu H, Gu J, Zhang Z, Dai J, Wei Y. Relation networks for object detection. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018:3588-3597.
16. Pawar SS, Prasanth Y. Multi-Objective Optimization Model for QoS-Enabled Web Service Selection in Service-Based Systems. New Review of Information Networking. 2017 Jan 2;22(1):34-53.
17. Phanikrishna C, Reddy AV. Contour tracking based knowledge extraction and object recognition using deep learning neural networks. In2016 2nd International Conference on Next Generation Computing Technologies (NGCT) 2016 Oct 14:352-354.
18. Sasikala N, Kishore PV. Train bogie part recognition with multi-object multi-template matching adaptive algorithm. Journal of King Saud University-Computer and Information Sciences. 2017 Oct 4.
19. Sreedevi, E., Prasanth, Y. A novel class balance ensemble classification model for application and object-oriented defect database. Journal of Advanced Research in Dynamical and Control Systems. 2017;9:702-26.
20. Rao, I. V. R., Anusha, S., Mohammad, A. B., Satish Kumar, D. Object tracking and object behavior recognition system in high dense crowd videos for video supervision: A review. Journal of Advanced Research in Dynamical and Control Systems. 2018;10(SP-2):377-80.
21. Sasikala N, Kishore PV, Prasad CR, Kumar EK, Kumar DA, Kumar MT, Prasad MV. Unifying Boundary, Region, Shape into Level Sets for Touching Object Segmentation in Train Rolling Stock High-Speed Video. IEEE Access. 2018 Oct 23;6:70368-77.
22. Bandi R, Amudhavel J. Object recognition using Keras with backend tensor flow. International Journal of Engineering and Technology (UAE). 2018;7(3.6):229-33.
23. Deepika, V., Rao, M. K., Kiranmai, N. Tokenization of news feed articles based on their similarity using machine learning techniques. Journal of Advanced Research in Dynamical and Control Systems. 2018;10(2):252-56.
24. Krishna Chaitanya, G., Meka, D. R., Vamsi, V. S., Ravi Karthik, M. V. S. A survey on Twitter sentimental analysis with machine learning techniques. International Journal of Engineering and Technology(UAE). 2018;7(2.32):462-65.
25. Anila, M., Pradeepini, G. Study of prediction algorithms for selecting appropriate classifiers in machine learning. Journal of Advanced Research in Dynamical and Control Systems. 2017;9(SP-18):257-68.
26. Venkata Naresh Mandhala, Debnath Bhattacharyya, Tai-hoon Kim. Hybrid Face Recognition using Image Feature Extraction: A Review. International Journal of Bio-Science and Bio-Technology, IJBSBT, August 2014;6(4):223-234.