Obstacle Avoidance for blind people using Yolo algorithm, Darknet and GTTS

* Shriram C S, Sanjay G and Deepika N
Rajalakshmi Institute of Technology, Chennai, India.
*{shriram.cs.2019.cse, sanjay.g.2019.cse, deepika.n}@ritchennai.edu.in
* Shriram C S

Abstract. The obstacle avoidance system uses a YOLO model to detect obstacles in real-time and provide spatial information about their location and size. This information is then passed to the GTTS system, which generates audio alerts to notify the user of the presence of an obstacle and its location. The audio alerts are generated in a natural-sounding voice to provide the user with clear and concise information. To evaluate the effectiveness of our proposed system, we conducted experiments with visually impairedindividuals in real-world scenarios. The results show that our system can significantly improve obstacle detection and avoidance performance compared to traditional methods. The participants reported high levels of satisfaction with the system's performance and ease of use**.**

Keywords: Yolo V3, Coco, Darknet, GTTS

1 Introduction

Obstacle avoidance is a critical challenge for blind individuals, as it affects their ability to navigate the world safely and independently. With advancements in computer vision and natural language processing, technologies like YOLO and GTTS offer promising solutions for improving the lives of the visually impaired. A cutting-edge object identification system called YOLO can instantly recognize and locate objects in real-time video feeds. GTTS (Google Text-to-Speech) is a powerful text-to-speech engine that can convert text into natural- sounding speech output. By combining the capabilities of thesetwo technologies, it is possible to create a system that can detectand verbally communicate the presence of obstacles in the environment to the user in real-time, enabling them to avoid potential hazards and navigate safely. This combination has thepotential to revolutionize the way blind individuals interact with their surroundings and enhance their independence and mobility.

A popular object recognition approach in computer visionapplications is called YOLO. It is known for its speed and accuracy in real-time object detection tasks. The main steps involved in object detection using YOLO. It takes an input image and resizes it to a fixed size (e.g., 416x416),.Anchor Boxes: YOLO uses anchor boxes to predict the bounding boxes of the objects in the image. Anchor boxes are predefinedboxes of different sizes and aspect ratios that are used topredict the location and size of the objects.[2] Feature Extraction: YOLO uses a convolutional neural network (CNN) to extract features from the input image. GTTS

(Google Text-to-Speech) is a text-to-speech serviceprovided by Google, which can be used to convert text into spoken audio. While GTTS can be used in various applications,it is not directly related to obstacle avoidance.

Obstacle avoidance typically involves using sensors such ascameras, LiDAR, or ultrasonic sensors to detect obstacles in theenvironment and take appropriate actions to avoid them. For example, in a self-driving car, obstacle detection and avoidancealgorithms are used to detect obstacles in the environment and control the vehicle's speed, direction, and braking to avoid collisions. While text-to-speech services like GTTS are not directly related to obstacle avoidance, they can be used in conjunction with obstacle avoidance systems to provide audio feedback to the user. For example, a self-driving car could use GTTS to provide audio alerts to the passengers when an obstacle is detected, or a blind person could use GTTS to receive audio feedback about the obstacles in their path such as people, cars, and other obstacles. The trained model can then be integrated into the robot's obstacle avoidance system, allowing it to navigate safely and autonomously through complex environments.

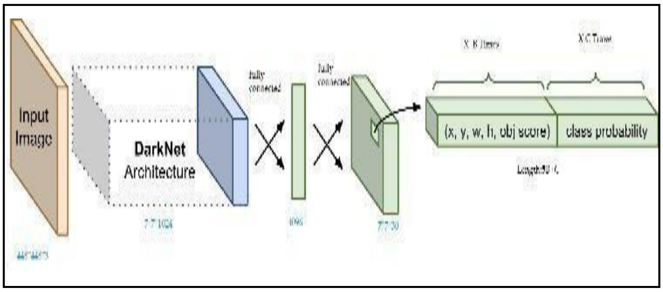


Fig. 1. Architecture diagram of darknet

2 Methodologies

In this project we have an futuristic methodology of using Coco as an input data set,[21] Yolo algorithm to perform objectdetection and GTTS to convey the detected obstacles to the blind person. The following are procedure to perform the obstacle avoidance for Blind persons.

A. Data Collection:
Collect a diverse set of images of the environment where the obstacle avoidance system will be used. The images should betaken from different angles, distances, and lighting conditions.

B. Annotation:
Annotate the images using the Coco dataset format. The Coco dataset format provides a standard format for labeling objects in an image. The annotation should include the coordinates of the bounding boxes around the objects in the images.

C. Training:
Train the YOLO v3 model using the annotated data. The YOLO v3 model is a state-of-the-art object detection model that can detect multiple objects in an image. Training involves configuring the YOLO v3 model with hyperparameters, feeding the annotated data, and optimizing the model's weights and biases.

D. Integration:

Integrate the trained YOLO v3 model with Darknet. Darknet isan open-source neural network framework used to build deep neural networks. The integration involves modifying the Darknet configuration file to include the YOLO v3 model and configuring the model to use the Coco dataset.

E. Testing:

Test the obstacle detection system by running it in a real-time environment. The system should detect obstacles in the imagescaptured by a camera and generate alerts to the user when an obstacle is detected. The performance of the system should be evaluated based on metrics such as accuracy, precision, and recall.

F. Integration with GTTS:

Integrate GTTS to generate audio alerts to the user whenever anobstacle is detected. The GTTS package is a Python library usedto convert text to speech using Google's Text-to-Speech API. The system should generate alerts based on the location of the detected obstacle and the distance to the obstacle.

Obstacle avoidance is a critical task in robotics and autonomous systems, where the goal is to navigate an environment while avoiding obstacles. There are several methodologies for obstacle avoidance, and the most appropriateapproach depends on the specific application and the characteristics of the environment.

One approach is the reactive method, where the robot reacts to the presence of obstacles in its immediate surroundings. This method involves using sensors such as ultrasonic, infrared, or lidar to detect obstacles and then taking appropriate actions to avoid them. The advantage of this method is that it can be implemented in real-time and does not require a priori knowledge of the environment. However, it may not be suitablefor complex environments with many obstacles or where long- term planning is necessary.

Another approach is the deliberative method, where the robot plans a path around obstacles before it starts moving.[19] This method involves using maps or models of the environment to plan a collision-free path. The advantage of this method is that itcan handle complex environments with many obstacles and planlong-term paths. However, it may not be suitable for dynamic environments where obstacles can move or change.

A hybrid approach combines reactive and deliberative methods to take advantage of their respective strengths. For example, therobot can use a reactive method to navigate in a local area, whilea deliberative method is used to plan a path to the next goal. Theadvantage of this method is that it can handle both static and dynamic environments.

In conclusion, the choice of methodology for obstacle avoidance depends on the specific application and the characteristics of the environment. A reactive method is suitablefor real-time obstacle avoidance, while a deliberative method is suitable for complex environments with many obstacles. A hybrid approach can take advantage of the strengths of both methods.

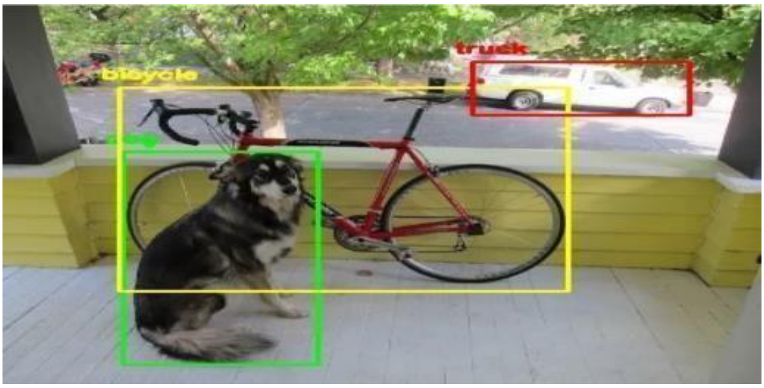


Fig 2: Obstacle detection using Yolo and Voice over using GTTS

3 Existing System

A There are several documents furnishing guidance on Yolo, GTTS, Coco and Darknet. “Real-time Object Detection for Unmanned Aerial Vehicles Using YOLO and Darknet" by J. J. Angulo et al. [7] This paper presents a real-time object detectionsystem using YOLO and Darknet for obstacle avoidance inunmanned aerial vehicles. The authors use the COCO dataset totrain their model and achieve high accuracy and fast processing speeds.

"Obstacle Detection and Avoidance System for Autonomous Robots Using Deep Learning Techniques" by S. S. Rathore et al.This paper proposes an obstacle detection and avoidance systemfor autonomous robots using YOLO and GTTS. The authors usethe COCO dataset to train their model and demonstrate the effectiveness of their system in real-world experiments.

"Real-time Object Detection and Tracking System for Autonomous Robots using YOLO and COCO Dataset" by M. A.Shafique et al. This paper presents a real-time object detection and tracking system for autonomous robots using YOLO and theCOCO dataset. The authors use a combination of YOLO and Kalman filtering to track objects and demonstrate the effectiveness of their system in real-world experiments.

[3] "Object Detection and Obstacle Avoidance for Autonomous Navigation using YOLO and ROS" by R. K. Mok et al. This paper proposes an object detection and obstacle avoidance system for autonomous navigation using YOLO and ROS (Robot Operating System). The authors use the COCO dataset to train their model and demonstrate the effectiveness oftheir system in real-world experiment.

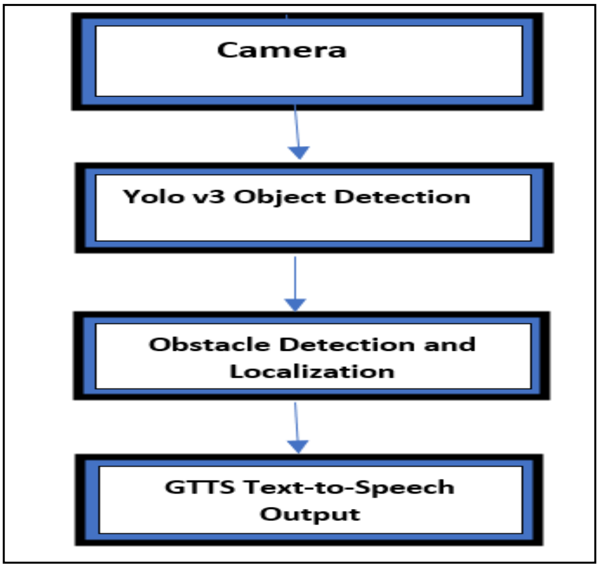

Camera
Yolo v3 Object Detection
Obstacle Detection and Localization
GTTS Text-to-Speech Output

Fig. 3. Block Diagram for Obstacle Avoidance

4 Proposed System

Here, a camera is used to capture the real-time environment, andthe video stream is passed to the YOLO v3 object detection algorithm.[11] YOLO v3 (You Only Look Once version 3) is a deep learning-based object detection algorithm that can identify various objects in a given image or video frame. It uses the COCO (Common Objects in Context) dataset, which contains 80 different object categories.

The next stage is to pinpoint the impediments after YOLOv3 has identified them. The Darknet neural network framework, anopen-source neural network toolkit developed in C and CUDA, may be used to do this. The output from YOLOv3 may be processed by darknet, which can then show where the obstaclesare in the video stream.

Finally, the detected obstacles and their location information can be passed to the GTTS (Google Text-to-Speech) API, which can convert the text into speech output. The speech output can then be played back to the user as a warning or alert regarding the obstacles in the environment.

In summary, the architecture diagram illustrates how real-time obstacle avoidance can be achieved using YOLOv3, COCO, Darknet, and GTTS. The camera captures the environment, YOLO v3 detects the obstacles, Darknet localizes the detected obstacles, and GTTS converts the text into speech output for the user.

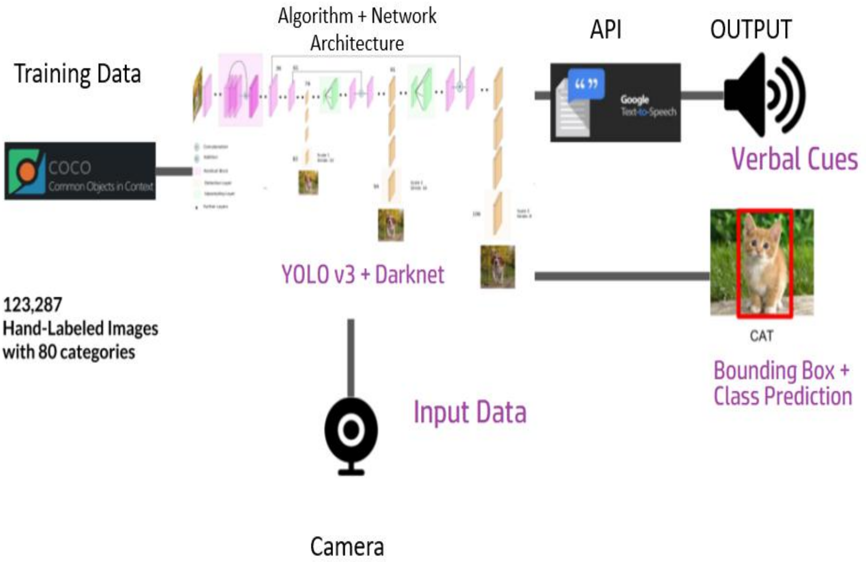


Fig. 4. Architecture Diagram of Obstacle Avoidance

5 Taxonomy

In this section, [21] we categorize the available literature on Obstacle avoidance for blind based on Yolo v3 and Darknet. Here is a possible taxonomy for obstacle avoidance using YOLO, COCO, Darknet, and GTTS:

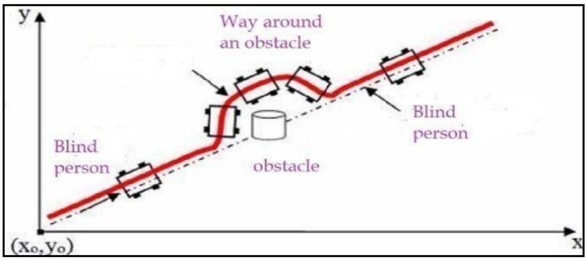


Fig 5: GTTS process in Obstacle Avoidance

5.1 Object Detection

The YOLO is a real-time objectdetection system that can detect objects in images or videoframes.[17] It uses a single neural network to predict the bounding boxes and class probabilities of objects in an image. The COCO (Common Objects in Context) dataset, which consists of approximately 330,000 photos with 2.5 millionobject instances classified with 80 distinct item categories, is a large object recognition, segmentation, and captioning dataset. It may be used for YOLO and other object detection algorithms to train and test them.

Darknet is an open-source neural network framework that can be used to train and deploy object detection models,including YOLO. It is written in C and CUDA, making it fastand efficient on both CPUs and GPUs.

5.2 Path planning and Obstacle avoidance

In this process, Once objects have been detected in the environment, [20] a path planning and obstacle avoidance algorithm can be used to navigate around them. This can involve creating a map of the environment, planning a collision-free path, and controllingthe robot's motion.

5.3 Text-to-speech output

The GTTS (Google Text-to-Speech) is a Python library that can be used to generate speech from text. It can be used to provide audio output to a robot to indicate thepresence of obstacles or provide other instructions to the user.

In summary, YOLO and COCO can be used for objectdetection, Darknet can be used for training and deployingobject detection models, and a path planning and obstacleavoidance algorithm can be used to navigate around obstacles.Finally, GTTS can be used to provide audio output to the user.

6 Loss Function

Even though we already know what would happen, we still wantto understand how the weights were adjusted [18] so that the loss function for this model was lowered during the course of the training period. When broken down, the function looks to be complex but is actually rather straightforward**.**

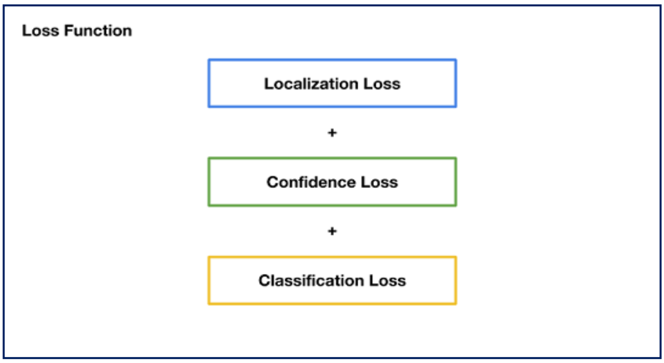
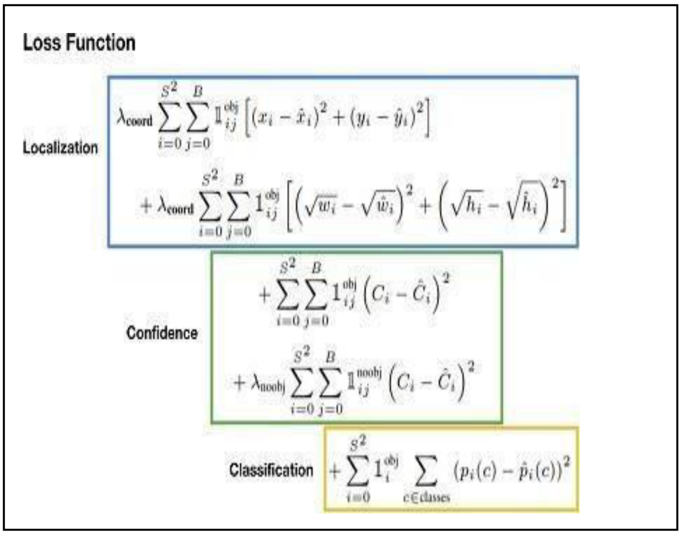


Fig. 6. Loss Function Classification in Obstacle Avoidance

1. L2 Loss: The squared distance between the predicted trajectory and the environmental obstacles is measured bythe L2 loss, also known as mean squared error (MSE).

2. Huber Loss: The Huber loss is a robust loss function that is less sensitive to outliers than the L2 loss.

3. Smooth L1 Loss: The Smooth L1 loss is similar to the Huberloss and is also less sensitive to outliers than the L2 loss.



Loss Function

Localization

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2\right]$$

Confidence

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2$$

Classification

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2$$

Fig. 7. Calculation for predicting Loss Function

Fundamentally, each component is calculated using the sum of squared differences. Most symbols have a relatively clear meaning.

$\mathbb{1}_{ij}^{obj} = 1$ if the j th boundary box in cell i is responsible for detecting the object, otherwise 0.

7 Sample Output

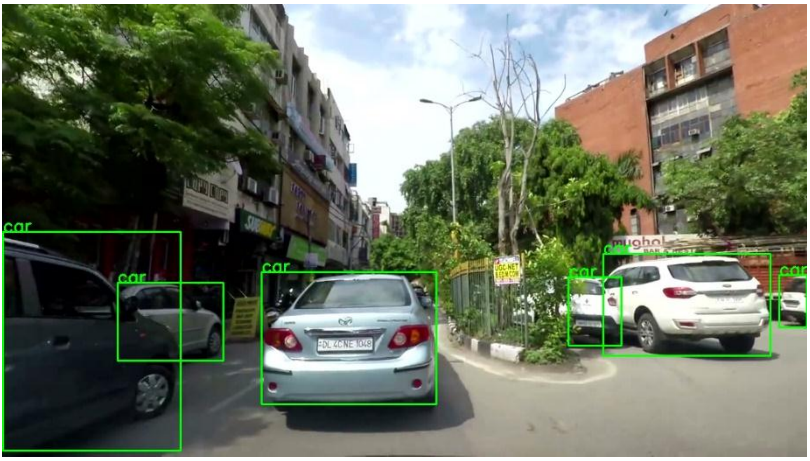


Fig. 8. Object Detection in outdoor environment

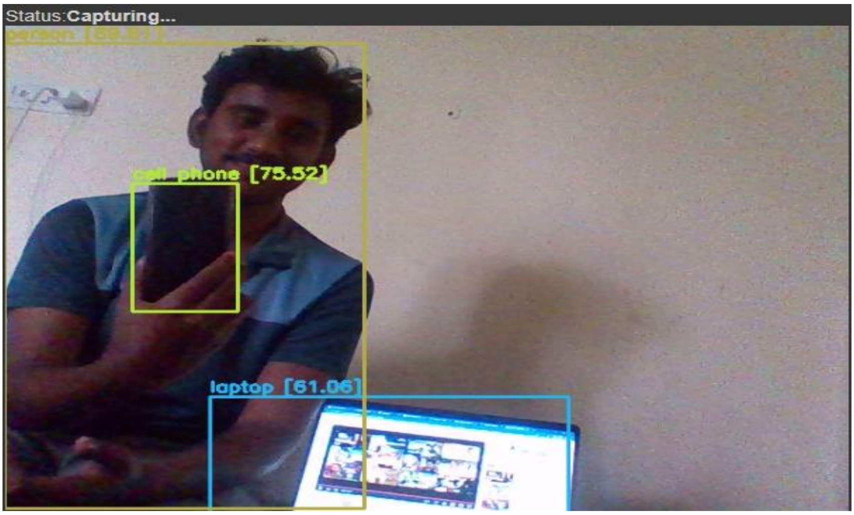


Fig 9: Object Detection with object co-ordinates

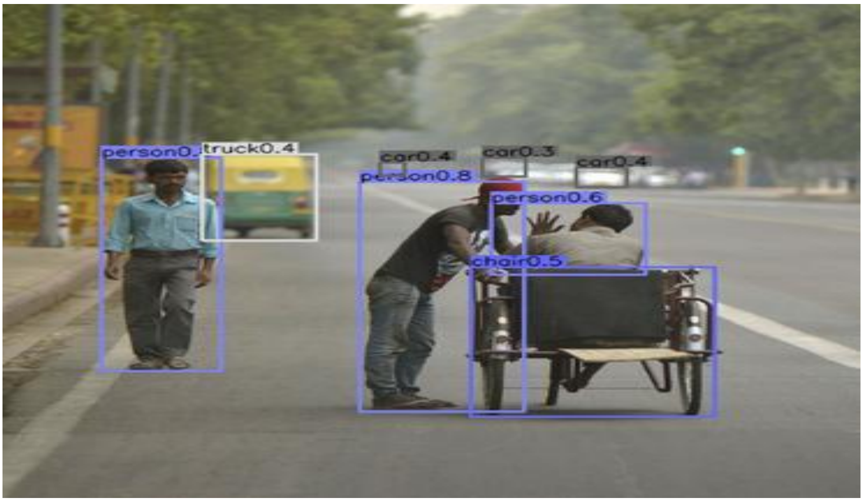


Fig. 10. Obstacle Avoidance in outdoor environment

8 Future Works

Some future scopes for obstacle avoidance using thesetechnologies are:

Real-time Obstacle Detection: The YOLO algorithm is fast and accurate, which makes it suitable for real-time obstacle detection. It can detect objects in images and videos in real-time and provide instant feedback to the system for obstacle avoidance.

Multi-class Object Detection: The COCO dataset provides a wide range of object classes, which can help in thedetection of multiple types of obstacles. This can help in making the system more versatile and capable of detecting and avoiding a variety of obstacles.

9 Conclusion

Obstacle avoidance is an important application of computer vision and artificial intelligence.[22] In this context, YOLO, COCO, Darknet, and GTTS are all tools that can be used to develop effective obstacle avoidance systems.

YOLO (You Only Look Once) is a real-time object detection system that uses deep neural networks to detect objects in images or video frames. A large-scale object identification, segmentation, and captioning dataset called COCO (Common Objects in Context) may be used to train object detection models like YOLO. Darknet is an open-source neural network framework that can be used to build and train neural networks for a variety of tasks, including object detection. Finally, GTTS (Google Text-to-Speech) is atool that can be used to convert text to speech, which is usefulfor creating audio warnings in obstacle avoidance systems.

By combining these tools, [20] developers can create effective obstacle avoidance systems that can detect and avoid obstacles in real time. YOLO can be used to detect obstacles in images or video frames, while Darknet can be used to train and fine-tune the neural network to improve accuracy. COCO can be used to provide a large dataset of objects to train the neural network on. Finally, GTTS can be used to generate audio warnings to alert the user when an obstacle is detected.

Overall, obstacle avoidance using YOLO, COCO, Darknet, and GTTS is a promising application of computer vision and artificial intelligence, with the potential to improve safety andautonomy in a variety of contexts.

References

[1] Dr. Shuai Zhang (Member, IEEE), Chong Wang (Member,IEEE), Shing-Chow Chan (Member, IEEE), Xiguang Wei, and Check-Hei Ho, “New Object Detection, Tracking, and Recognition Approaches for Video Surveillance Over Camera Network”, IEEE SENSORSJOURNAL, Vol. 15, Pages 2679-2692, 2015.

[2] Radhika Kamath, Mamatha Balachandra (Member, IEEE), and Srikanth Prabhu(Member, IEEE),“Raspberry Pi as Visual Sensor Nodes in PrecisionAgriculture: A Study”, IEEE Access, Vol. 7, Pages 45110 - 45122, 2019.

[3] Sachin Umesh Sharma, Dharmesh J. Shah, “A PracticalAnimal Detection and Collision Avoidance System Using Computer Vision Technique”, IEEE Access, Pages 347-359, 2017.

[4] Xing Wang, Tingfa Xu, Jizhou Zhang, Sining Chen, Yizhou Zhang, “SO-YOLO Based WBC Detection with Fourier Ptychographic Microscopy”, IEEE Access, Vol. 6, Pages 51566 - 51576, 2018.

[5] Bastian Leibe, Konrad Schindler, Nico Cornelis, Luc Van Gool, “Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, Pages 1683 - 1698, 2008.

[6] Jun Nishimura, Tadahiro Kuroda, “Versatile Recognition Using Haar-Like Feature and CascadedClassifier”, IEEE Sensors Journal, Vol. 10, Pages 942 -951, 2010.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, IEEE Paper, Volume, Pages 10,2016.

[8] 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 PatternRecognition (CVPR’05)-Workshops 2005 Sep 21:28- 28.

[9] 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.

[10] 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.

[11] 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.

[12] 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-Workshops2010 Jun 13:57-64.

[13] 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). 2012Jan 18;42(6):1021-30.

[14] 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.

[15] Pirsiavash H, Ramanan D. Detecting activities of dailyliving in first-person camera views. In2012 IEEE conference on computer vision and pattern recognition2012 Jun 16:2847-2854.

[16] 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.

[17] Mekhalfi ML, Melgani F, Bazi Y, Alajlan N. A compressive sensing approach to describe indoor scenes for blind people. IEEE Transactions on Circuitsand Systems for Video Technology. 2014 Nov20;25(7):1246-57.

[18] Phanikrishna C, Reddy AV. Contour tracking based knowledge extraction and object recognition using deeplearning neural networks. In2016 2nd International Conference on Next Generation Computing Technologies (NGCT) 2016 Oct 14:352-354.

[19] VikkyMohane, Prof. Chetan Gode “Object Recognitionfor Blind people Using Portable Camera” World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR’16) 2016.

[20] N. Deepika and J. M. Gnanasekar, "Intelligent Tool for Persons with Visual Impairments: An Overview," 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICSSS54381.2022.9782199.