Object Detection for Blind

2232-CSE-6367-001-COMPUTER VISION

Anil Bhavanam Honnappa Gowda - 1001966285

***Abstract*—Yolov3 is an advanced real-time object detection algorithm that achieves high accuracy and fast processing speed. The algorithm is based on deep neural networks and uses a single convolutional neural network to predict the bounding boxes and class probabilities of objects in an image. This makes the algorithm highly efficient and effective for real-time object detection in applications such as surveillance, autonomous vehicles, and robotics. Yolov3 can detect over 80 different object categories and can accurately detect objects even in crowded and complex scenes. Its high accuracy and efficiency make it one of the most popular and widely used object detection algorithms in the field of computer vision. In addition, Yolov3 supports transfer learning, where a pre-trained model can be fine-tuned for a specific task, reducing the need for large datasets and improving the detection accuracy. The versatility and flexibility of Yolov3 make it a powerful tool for various real-world applications that require reliable and efficient object detection.**

***Index Terms*—Yolov3, object detection, convolutional neural network.**

I. Introduction

Many people all across the world have vision problems, which makes it challenging for them to find their possessions and get around on their own. This project seeks to identify potential barriers that blind or visually impaired people might run against and to offer modest support. Through a camera input, it employs object detection to detect impediments in real-time. A verbal alarm is produced when an object is found to direct the user around the obstruction. To do this, the yolov3 object detection paradigm is used.

Computer vision is a branch of AI and machine learning that focuses on giving computers the ability to interpret and comprehend visual data from their surroundings. In order for computers to evaluate, interpret, and extract valuable information from digital photos and videos, algorithms and models must be developed. Numerous businesses, including healthcare, transportation, security, and entertainment, use computer vision in diverse ways. Object identification, facial recognition, autonomous cars, and medical imaging are a few applications of computer vision. Computer vision is a growing field with many interesting future prospects.

A key component of computer vision, object detection identifies the type and placement of objects in an image or video. With the advent of deep learning-based models like YOLO (You Only Look Once) and Faster R-CNN (Region-based Convolutional Neural Networks), object identification has grown in significance in recent years. Even in complex scenarios with numerous objects and occlusions, these models can accurately and effectively recognize objects in real-time. Applications for object detection are numerous, ranging from robots and self-driving cars to surveillance systems and imaging in the medical field. Object detection is used in the healthcare industry to find anomalies in medical imaging. Among many other uses, the technology is employed in activity recognition and pedestrian detection. Overall,In general, object detection is a critical technology that has the capacity to completely transform a variety of fields and sectors.

II. Related Work

A. Literature survey

To enable fast and accurate detection and classification of underwater targets during the operation of intelligent underwater robots, a modified version of YOLOv3 algorithm, called YOLOv3-UW, is proposed. This algorithm improves the clustering algorithm of input datasets, streamlines the network architecture, and enhances the residual module to achieve better detection accuracy and speed than the original YOLOv3 algorithm. The experimental results demonstrate the superiority of YOLOv3-UW in terms of detection accuracy and speed [1].

Improved YOLOv3-based neural network for object de-identification in surveillance systems. The proposed network reduces the number of feature extractors from 53 to 24 and the number of box detection parts from 53 to 15, resulting in a significant increase in detection speed while maintaining comparable detection accuracy. The proposed algorithm is tested on the WIDER FACE dataset and an additional dataset, and the results demonstrate that it outperforms YOLOv3-tiny and achieves a faster detection speed than YOLOv3 [2].

A novel vehicle detection algorithm, ADE-YOLOV3, is proposed to address the issue of repeated detection in YOLOV3 algorithm. This algorithm utilizes K-means clustering to determine the number of target frames and aspect ratio based on the width and height values of the vehicle. The anchor parameters are then reset based on the results obtained by clustering to improve the pertinence of the ADE-YOLOV3 network in vehicle detection. The migration learning method is used to optimize the network structure and obtain the optimal weight model, which improves the training precision of the model. The experimental results show that ADE-YOLOV3 outperforms the original YOLOV3 method in terms of mAP, repeated detection rate, and detection speed, while effectively avoiding the problem of repeated detection [3].

III. Existing System.

The world has a significant population of people who are fully disabled and face various challenges while running. The steering task relies heavily on visual data, which puts visually impaired individuals at a disadvantage because they lack essential information about their surroundings. With the progress of comprehensive technology, it has become possible to enhance the support for visually impaired individuals and improve the quality of life for people with disabilities. Various aids are now available for individuals with visual impairment.

A. Object detection for visually impaired people

The application utilizes Android platform and multiple supporting libraries. It captures an image of the surroundings using the camera of an Android smartphone, which is then stored in the device's memory. The image is processed using libraries such as openCV and Google Cloud API. The Google Cloud Vision API is employed to compare the input image with millions of other images present in the COCO dataset. The processed image is sent to the cloud via the internet. Eventually, the application identifies the objects present in the user's surroundings and provides an audio output.

B. Object recognition for blind people based on features extraction

The application has several distinct goals that it aims to achieve. One of the main objectives is object detection, where it identifies everyday objects and provides an auditory output for the user. Additionally, the application focuses on obstacle detection by using sound alerts to signal the presence of walls, tables, doors, and other obstacles. To determine the distance between the user and the obstacles, sensors are utilized.

To achieve these objectives, the application employs various image processing tools. First, the frames of a video are matched with objects stored in the database by computing key points and extracting features to make a matching between the key points. For subsequent frames, they are matched with previous frames to predict whether they are the same or not. If different, the application computes keywords and extracts them to correspond with those extracted from objects stored in the database.

C. White cane

The white cane is a tool commonly used by individuals who are blind or have visual impairments. It enables them to detect obstacles and orient themselves in their surroundings, as well as aids in navigation. There are various types of white canes, including long canes, guide canes, identification canes, and support canes.

D. R-CNN

Ross Girshick and his team proposed a technique to overcome the challenge of dealing with a large number of regions in an image. They introduced the concept of "region proposals" where only 2000 regions are selected from an image for analysis, instead of processing the entire image. These region proposals are generated using the selective search algorithm, which involves generating many candidate regions, combining similar ones, and producing final proposals. The development of better neural networks is crucial for the advancement of target detection algorithms. Object detection systems based on convolutional neural networks (CNN) can be classified into two types: two-phase identifier type (such as R-CNN, R-FCN, and FPN) and single-stage indicator type (such as YOLO, SSD, and RetinaNet). The former produces a set of candidate regions as input for CNN, while the latter directly solves the object detection problem as a regression task without generating candidate regions.

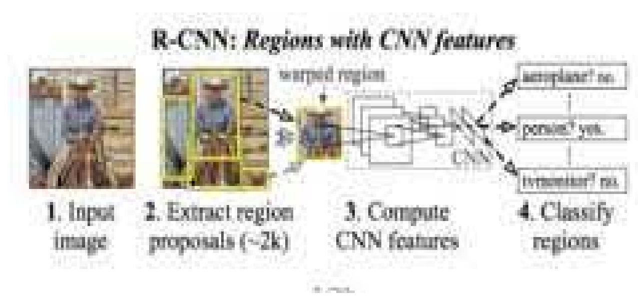

Fig. 1. R-CNN With Features

// Refference: tacken from google images

E. Fast R-CNN

The author of the previous R-CNN paper addressed some of its drawbacks and developed a faster object detection algorithm called Fast R-CNN. This algorithm is similar to R-CNN, but instead of feeding the CNN with region proposals, it feeds the input image to generate a convolutional feature map. The feature map is then used to identify the regions of the proposals and warp them into squares. A RoI pooling layer is used to reshape them into a fixed size that can be fed into a fully connected layer.

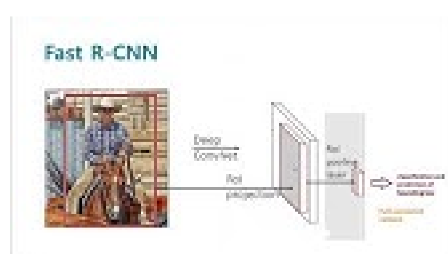

Fig. 2. Fast R-CNN Features

// Refference: tacken from google images

From the RoI feature vector, a softmax layer is used to predict the class of the proposed region and the offset values for the bounding box.

F. YOLO — You Only Look Once

Previous object detection algorithms used region-based methods to localize objects in an image. These methods only looked at parts of the image that were likely to contain the object, rather than the entire image. YOLO, or You Only Look Once, is a different type of object detection algorithm that uses a single convolutional neural network to predict both the bounding boxes and class probabilities for objects in the image.

YOLO divides an image into a grid of SxS cells, and for each cell, it predicts m bounding boxes. For each bounding box, the network outputs a class probability and offset values. Objects in the image are detected by selecting bounding boxes with class probabilities above a threshold.

Compared to other object detection algorithms, YOLO is much faster. However, it struggles with detecting small objects, such as a group of birds, due to the spatial limitations of the algorithm.

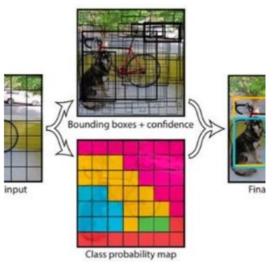

Fig. 3. Yolo

// Refference: tacken from google images

IV. Proposed Methodology.

Object detection is indeed a powerful technique that has a wide range of applications in various fields such as autonomous vehicles, surveillance, robotics, and more. In the context of this project, YOLOv3, a state-of-the-art object detection algorithm, is used to detect obstacles for blind individuals.

A. Problem statement

Helping the visually challenged people by informing the obstacle through object detection using yolov3.

B. Objectives

- The primary objective of object detection is to accurately detect and classify objects within an image. YOLOv3 aims to achieve this by using a deep neural network that can recognize and categorize objects with high accuracy.
- Providing real-time object detection to give blind individuals a better sense of their surroundings and enable them to find there objects.

C. Dataset: Coco dataset:

The dataset used in YOLOv3 is the Common Objects in Context (COCO) dataset. It is a large-scale object detection, segmentation, and captioning dataset that contains more than 330,000 images and over 2.5 million object instances. The dataset contains 80 different object categories, including people, animals, vehicles, furniture, and more. The images in the dataset are annotated with bounding boxes that identify the location of each object in the image. This dataset is commonly used for training and evaluating object detection models, including YOLOv3.

D. Libraries Used.

- OpenCV, or Open Source Computer Vision Library, is a software library that's free and open-source. It's used for computer vision and machine learning, and is popular in a variety of fields, including robotics, surveillance, automotive, and medical imaging. The library has a large community of developers, researchers, and users who work together to maintain and improve it. In this project we will use openCV to process the input.
- numpy : NumPy is a Python library used for scientific computing and data analysis. It provides support for multidimensional arrays and matrices, which can be used to store and manipulate large datasets efficiently. NumPy is essential for working with data in Python, as it provides a fast and efficient way to perform mathematical operations on arrays, such as adding, subtracting, multiplying, and dividing. NumPy also provides a number of functions for statistical analysis, linear algebra, and Fourier analysis. Overall, NumPy is a powerful library that is widely used in scientific computing, data analysis, and machine learning applications.
- pyttsx3 : pyttsx3 is a Python module that allows for the conversion of written text into spoken words using text-to-speech engines installed on the user's device. With a straightforward API, it provides support for generating speech in different languages and voices. The library is commonly utilized in software applications that require speech output such as assistive technology, virtual assistants, and educational software. Pyttsx3 can work with various text-to-speech engines including sapi5, dummy, and espeak.
- time : Time library that enables us to quantify and measure the duration of events and the intervals between them.
- pywin32 : To run pyttsx3 library successfully we need pywin32 package to be installed and we can install pywin32 only in windows but not in MAC.

E. Working procedure:

After running the code, the surveillance camera will open and the input is taken from the webcam and then object is detected in real time using yolov3 and the output, object detected is shown in the surveillance camera with object name and voice alert or audio alert is given with name of the object detected.

F. Object detection with Yolo model:

The training data for the object detection model is obtained from the Common Objects in Context (COCO) dataset, which includes images and their corresponding labels. The model used in this project is based on the You Only Look Once (YOLO) algorithm, which employs a complex Convolutional Neural Network (CNN) architecture called the Darknet. Despite the fact that an optimized and more sophisticated YOLOv3 model is used, the basic principles of the original YOLO algorithm are explained. The Darknet is set up using the cv2 package in Python, using the yolov3.cfg configuration file. The input for the model is captured from the webcam.

G. Working process of yolov3:

The YOLOv3 algorithm processes images by dividing them into a 13x13 grid of cells, with the size of each cell dependent on the input image size. Each cell predicts multiple bounding boxes for potential objects, along with the confidence level of the box accurately enclosing an object and the probability of the enclosed object belonging to a specific class. However, many of these bounding boxes are filtered out due to low confidence scores or overlapping with other highly confident bounding boxes of the same object. This technique is called non-maximum suppression.

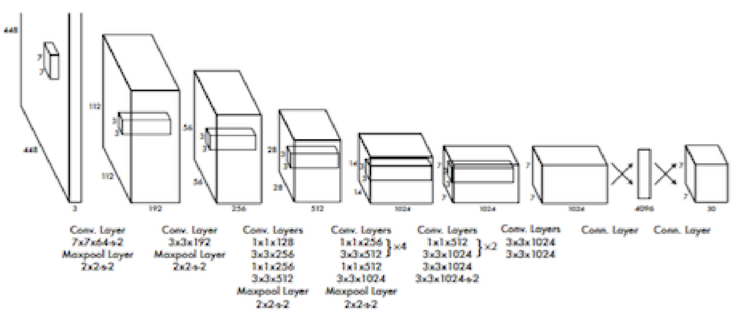

Fig. 4. Yolo

// Refference: tacken from google images

The remaining boxes undergo further processing, where only those above a certain confidence threshold are considered. Non-maximum suppression is then applied again to remove redundant boxes based on a specified threshold, and the resulting matched objects are given as output according to the datasets used.

V. Results:

A. Obeject Detection

By taking the input from the webcam object detection will be perform, since this is to assist the blind people we are doing real time object detection.

In the image the, Cell phone and person is detected at the time of implementation and the voice alert is given that cell

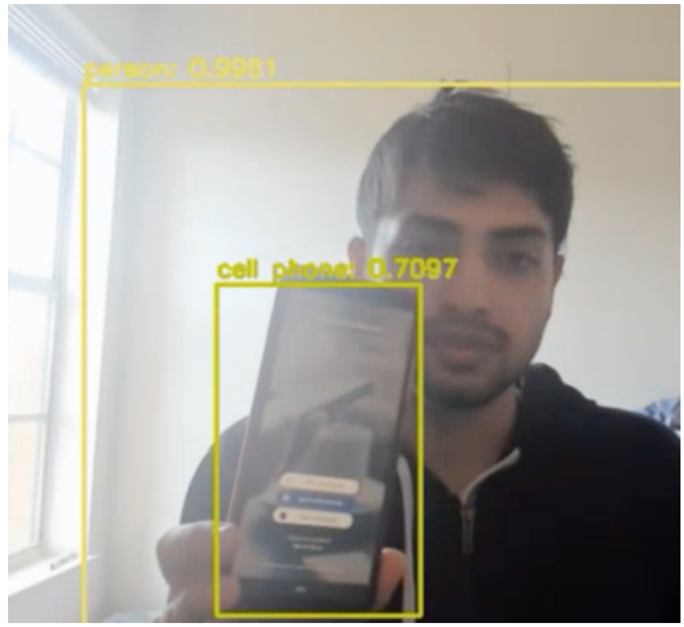

Fig. 5. result image 1

// screenshot of implementation

is moving and person is moving along with which direction the object is moving.

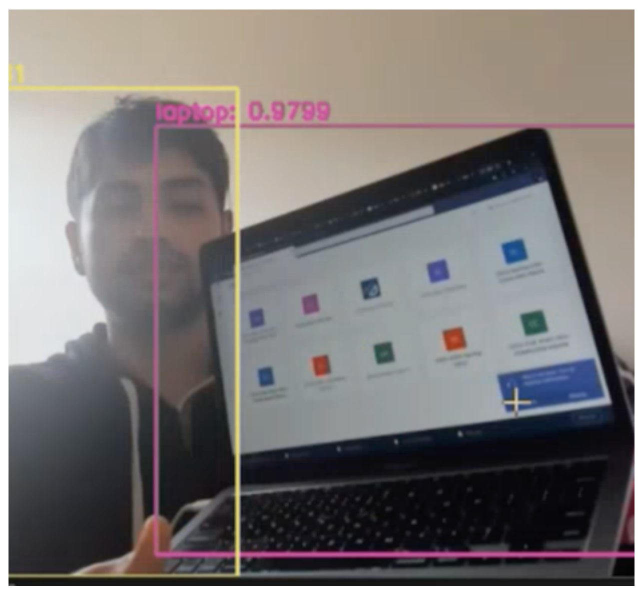

Fig. 6. result image 2

// screenshot of implementation

In the image the, Laptop and person is detected at the time of implementation and the voice alert is given that cell is moving and person is moving along with which direction the object is moving.

In the above image you can see the bottle detection along with the person .

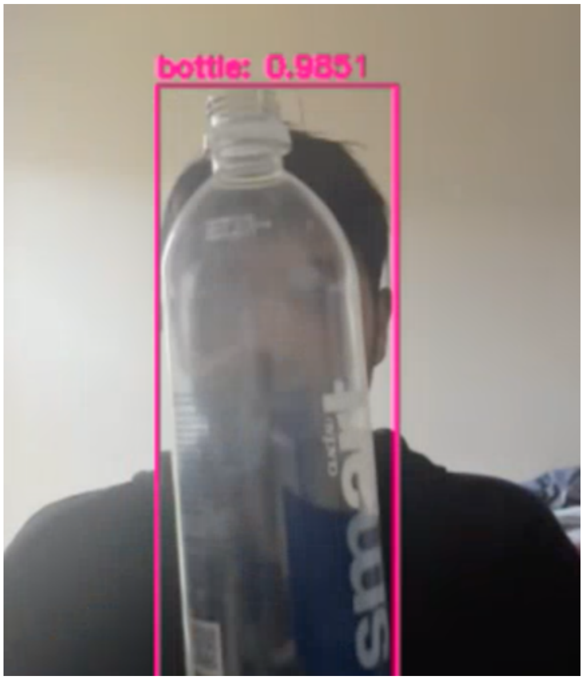

Fig. 7. result image 3

// screenshot of implementation

VI. Concusion:

Over the years, there have been some tools developed to assist visually impaired individuals in identifying common objects in their surroundings. However, only a small number of these tools are currently being used on a daily basis. Our objective is to introduce a visual substitution system that not only aids in object recognition but also in mobility. This system performs realtime object detection using webcam input, opencv , numpy , pyttsx3 , pywin32 yolov3 to detect objects in the environment of visually impaired individuals.

VII. Future Extensions:

In this project we can only detect the minimum number of objects , and in future the wall detection and obstacle detecting could also implement with this and also now yolov4 , yolov5, and yolov6 also available and we can also perform this project on those and we can use large number and variety of data. yolov3 could focus on improving the model's performance on small objects by incorporating advanced feature extraction techniques or optimizing the anchor box size and shape. And yolov3 could focus on incorporating advanced object tracking techniques to enable more reliable and accurate tracking of objects

References

For this i took four references, and the references are given bellow.

References

[1] Liu, Y., Zhang, G., and Zhang, Y. (2019, November). Vehicle detection method based on ADE-YOLOV3 algorithm. In 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (pp. 227-232). IEEE.

[2] Shi, T., Niu, Y., Liu, M., Yang, Y., Wang, C., Huang, Y. (2019, December). Underwater Dense Targets Detection and Classification based on YOLOv3. In 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO) (pp. 2595- 2600). IEEE.

[3] Lee, H., Kang, D., and Moon, W. (2007, June). Design and fabrication of the high directional ultrasonic ranging sensor to enhance the spatial resolution. In TRANSDUCERS 2007-2007 International Solid-State Sensors, Actuators and Microsystems Conference (pp. 1303-1306). IEEE.

[4] Heba Najm, Khirallah Elferjani and Alhaam Alariyibi. "Faster and Scalable MPI Applications Launching." IEEE Transactions on Parallel and Distributed Systems, 2022, pp. 1–17. IEEE Xplore, https://doi.org/10.1109/TPDS.2022.3218077

[5] ook IEEE Conference Template from the overleafe.com ,, https://www.overleaf.com/project/6457c399f96671675ed8e763 .