International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 04 | Apr 2023 www.irjet.net p-ISSN: 2395-0072

Object Detection and Tracking AI Robot

Jayati Bhardwaj[1], Mitali[2], Manu Verma[3], Madhav[4]

[1,2,3,4] *Department of CSE MIT, Moradabad, U.P, India*

Abstract: The task of object detection is essential in the fields of robotics and computer vision. This paper's goal is to provide an overview of recent developments in object detection utilizing AI robots. The study explores several object detection techniques, including deep learning-based methods and their drawbacks. The study also gives a general overview of how object detection is used in practical contexts like robots and self-driving cars. The advantages of deploying AI robots for object detection over conventional computer vision techniques are the main topics of discussion.

I. INTRODUCTION

Artificial intelligence (AI) and its uses in robotics have attracted increasing attention in recent years. Providing robots with the ability to detect and recognize items in their environment is one of the key issues in robotics. As the name suggests, object detection is the procedure of locating items in an image or video frame and creating bounding boxes around them[3]. For robots to carry out a variety of activities, including grabbing, manipulating, and navigating, this task is essential [7].

Traditionally, hand-crafted features like SIFT and HOG and machine learning methods like SVM and Random Forest have been used for object detection. However, these techniques have limits in terms of effectiveness and precision, particularly Growing interest has been shown in artificial intelligence (AI) and its uses in robotics in recent years [5]. Giving robots the ability to detect and recognize items in their environment is one of the most significant robotics challenges. The act of recognizing objects in an image or video frame and creating bounding boxes around them is known as object detection, as the name suggests. Robots must complete this activity in order to carry out other activities including grasping, manipulating, and navigating [9].

Traditionally, hand-crafted features and machine learning algorithms like SVM and Random Forest have been used for object detection [14]. Examples of these features include SIFT and HOG. However, these techniques have shortcomings, particularly in terms of efficiency and accuracy.

In this paper, we present a comprehensive examination of the state-of-the-art in object detection utilizing AI robots. We concentrate on recent advances in the area, discussing both conventional and deep learning-based approaches and highlighting their advantages and disadvantages [15]. We also go over numerous robotics uses for object detection, like grasping, manipulating, and navigation, and we give some insight into the difficulties and potential future directions of this field of study.

The goals of this paper are to offer an overview of the subject's current state and to encourage more research and growth in it

II. LITERATURE REVIEW

For many years, object identification algorithms have been being developed for AI robots. Early object detection techniques relied on manually made features, such as scale-invariant feature transforms, histograms of directed gradients, and sped-up robust features (SURF) (HOG)[6]. These methods have been widely applied to AI robots for object detection; however they have a number of shortcomings. For example, they are sensitive to the size and orientation of objects and have trouble capturing intricate shapes and textures.

The use of deep learning-based object detection techniques has increased recently[9]. These convolutional neural network (CNN)-based techniques have been shown to outperform more traditional feature-based methods in a variety of object identification tasks. The three most prevalent deep learning-based object detection techniques are region-based convolutional neural networks, You Only Look Once (YOLO), and single shot multi-box detectors (SSD) (R-CNN).These techniques may automatically learn object features and carry out object detection at the same time since they are trained from beginning to end.

Dealing with occlusions, where objects are partially or fully obscured from vision, is one of the difficulties in object detection. Several approaches have been put forth to deal with this problem, such as attention-based approaches, where the AI robot focuses on the portions of the image that are the most pertinent, and occlusion-

aware approaches, where the AI robot takes the occlusions into account when performing object detection.

[1] X.Zhang,Y Yang et.al. described a technique for object detection and tracking in outdoor settings that makes use of a mobile robot with a camera and a LiDAR sensor. The suggested approach combines a Kalman filter for object tracking with a deep neural network for detection.

J.Redmon and S.Divvala[2] work proposed a moving vehicle: Real-time Multiple Object Detection and Tracking The real-time object recognition system YOLO (You Only Look Once), which is described in this paper, can identify and track numerous objects in a video stream from a moving vehicle. The suggested approach concurrently detects and tracks objects in a video stream using a single neural network.

S.Wang,R.Clark and H.Wen[3] developed a real-time object identification and tracking system for autonomous driving applications is presented in this study. The suggested system combines the Kalman filter and Hungarian algorithm for object tracking with a deep neural network for object detection. The system is appropriate for real-time applications because it is built to operate with low-latency input and output. Table1 represents the comparative work analysis among different detection & tracking systems.

III.PROPOSED WORK

A. Algorithm Used:

Fast R-CNN enhances the object identification speed and precision of the original R-CNN method. Using a single deep neural network to carry out both object detection and feature extraction, as opposed to using several networks for these tasks, is the main novelty of Fast R-CNN. Fast R-CNN can now be both more accurate and faster than R-CNN thanks to this method.

Using a selective search technique, the Fast R-CNN algorithm first chooses a set of object proposals (i.e., areas in the image that could contain an object). Following the extraction of features from these suggestions using a convolutional neural network (CNN), the features are input into a set of fully connected layers that carry out the actual object classification and bounding box regression.

Overall, Fast R-CNN is a popular and successful object identification technique that has been employed in many fields, such as robotics, self-driving automobiles, and medical imaging.We used fast R-CNN because it has high accuracy among all algorithms. Table2 represents comparison among different object detection algorithms along with their accuracies.

Table 2: Comparison among algorithms accuracies used in Object Detection

Algorithm Used	Accuracy
ResNet	78%
R-CNN	81.71%
Faster R-CNN	84.9%
SSD	74%
YOLO	72.81%
HOG	82.1%

B.Materials Used

Hardware Components

1. Esp-32 Cam
2. Gear Motor
3. Wheels
4. Servo Motor
5. Portable Power Bank
6. Plastic Case
7. Arduino Nano
8. Hco5 Bluetooth Module

Software Used

1. Ardiuno Ide
2. Esp 32 Ai Camera
3. Ardiuno Automation

C. Working Model

The proposed method has been evaluated in three different types of scenarios for item detection and recognition in real-world settings. First, a semi-structured scene was taken into consideration in order to conduct a methodical analysis of the effectiveness depending on several factors. The second experiment featured two actual, chaotic settings. In this experiment the object had to found among variety of commonplace things like books, clocks, calendar and pens. Lastly, a picture dataset has been used to assess the system's performance through

object instance recognition and in comparison to other cutting-edge methods

An execution time performance analysis is offered as a conclusion. Two multi-jointed limbs and a humanoid torso equipped with a Microsoft TO40 pan-tilt-divergence stereo head were employed in the first two studies. Two Imaging Source DFK 31BF03-Z2 cameras mounted on the head take 1024x768-pixel colour images at 30 frames per second. High-resolution optical encoders give the motor positions, and the distance between the cameras is 270.

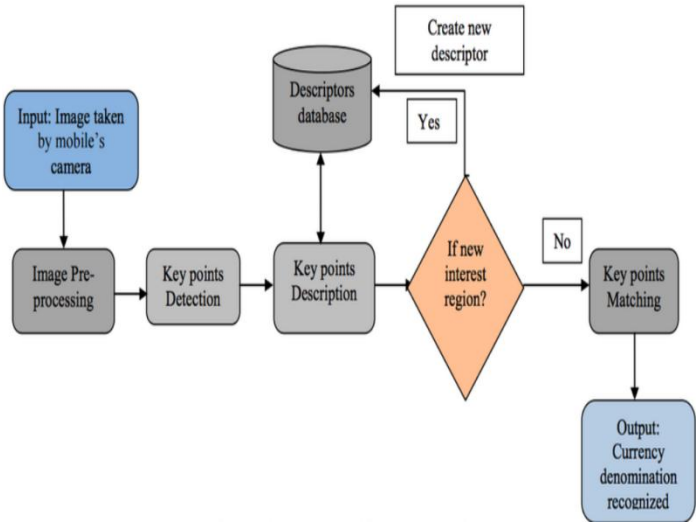

Fig.1 Object Detection Process

Figure shows flow chart of object detection. Firstly, input image will be taken and then it is pre-processed through key points .if new region is present then it will go to database otherwise output will be generated through matching of key points.

EXPERIMENT 1

In this experiment, the machine was positioned in front of tables that contained the some objects. In this experimental setup, the table was initially empty. However, after a short while, a human began placing and removing various objects from the table without directly interacting with the robot system. In this way, the motion cue assisted in determining the presence of a human in the robot workplace as well as the brand-new object instance on the table. Indeed, the three visual cues are given identical weight when segmentation results are calculated in this experiment. Targets have included various objects such as a red ball, a toy car, a bottle, and a money box.

The object's position and orientation were altered on every frame. The number of observed orientations clearly varies depending on the object in question; for instance, the toy vehicle was observed in 12 different orientations, whilst the red ball has only one orientation (roughly every 30 degrees). The accepted strategy starts with the image's capturing. This image serves as the input for two separate processes:

Segmenting the colour cue and the two other cues that were observed (i.e. motion and shape). The goal of this difference was to increase effectiveness. In order to segment the image and represent it in L1L2L3 coordinates, memory data on the various elements to be located is used. The other objects were also tested in similar ways. Figure 8 shows a few of the outcomes (just the final outcome).

It should be emphasised that the results are from a single study because the data are not random. As can be seen, only one object is searched at a time.

This result was reached after testing the system's functionality under a variety of circumstances that could lead to problems (such as shadows, flickering light sources, different light reflexes, partially visible objects, etc.). As shown, even when items changed their orientation, location within the picture, or angle of view from the cameras, all of the objects were still accurately identified.

EXPERIMENT 2

The things that were to be found and identified in this experiment were set up on a desk. Two unstructured environments were employed, each with a different set of commonplace items like textured books, pens, clocks, etc. Throughout the scenario under consideration, these objects were situated in various positions and/or orientations, which in some circumstances led to partial occlusion.

The motion cue once again causes a visual attention focus since, similar to the previous instance, a human is continuously interacting with the target objects but not with the robot system. The other two visual cues are required to distinguish between the target objects and other moving elements in the scene, such as the person. So, throughout the object recognition process, the three cues are equally important. In the first experiment, three different items—a toy car, a stapler, and a wooden cylinder—were used.

Notwithstanding the environment's features and those of the items themselves—including the toy automobile, whose colour was strikingly similar to that of its surroundings—all targets were accurately identified. The car and the stapler have been recognised and effectively identified in a case where two objects were found in a single photograph. This is similar to how the newly developed approach successfully focuses on the target object.

EXPERIMENT 3

In the final validation experiment, we use a public picture repository to compare the performance of our methodology against leading-edge techniques. Actually, there are many public image repositories available because object recognition is essential for many applications. These datasets give researchers the opportunity to assess their methods with a variety of objects and settings, as well as to assess how well they perform in comparison to other cutting-edge methods. These repositories could be categorised, nevertheless, according to the objectives they must achieve.

so the term "object recognition" might relate to a variety of application scenarios or it may be based on a particular set of input data. There are various levels of semantics (for example, category recognition, instance recognition, pose recognition, etc.). The required evaluation dataset must therefore comply with the demands of a particular technique. This dataset consists of thousands of RGB-D camera images of 300 common objects taken from different angles in household and business environments.

Because objects are organised into a hierarchy of 51 categories, each of which comprises three to fourteen instances, each object can only belong to one category To fully evaluate the segmentation procedure, ground truth photos are also provided. As a result, this image dataset enables the evaluation of object recognition methods on two different levels: •Level of categories. The process of categorising newly unseen objects based on previously seen objects from the same category is known as category recognition. In other words, this recognition level equates to determining if an object is an apple or a cup. Example level.

Is this Ester's or Angel's coffee cup? is the question that needs to be answered in this situation. Although the capacity to recognise objects at both levels is crucial for robotic tasks, only instance identification is taken into account in this work because no category abstraction was done. Finding the specific physical instance of an object that has already been presented is the goal of the recognition algorithm.

IV.RESULTS & CONCLUSION

After implementation of above proposed methods, we find the faster R-CNN having highest accuracy. So, we implemented the algorithm for object detection. For the control we use voice command implement with the help of HC05 Bluetooth module and for the manual control we use esp-32 cam connected with Arduino IDE. Figure 2 represents the final results of object detection by the proposed method.

Modules we implemented successfully

1. Detection
2. Tracking
3. Movement detection
4. Lane tracking
5. Avoid Obstacles
6. Interaction with humans

With the help of previously done research we are able to achieve all the modules with good accuracy. Combined accuracy of all the modules is nearly 76%.

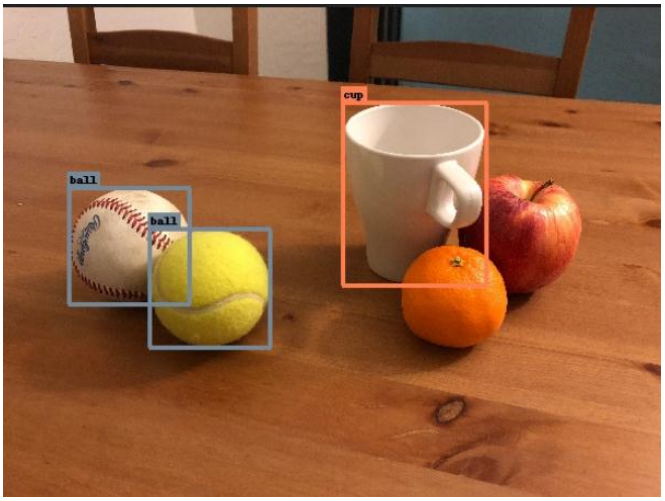

Fig.2 System Detecting Objects

Table1: Comparison among some object detection and tracking systems

	X.Zhang,Y Yang, and Y.Liu [1]	J.Redmon and S.Divvala[2]	S.Wang,R.Clark and H.Wen[3]
Title	"Object detection and tracking using a Mobile Robot"	"Real Time Tracking while driving a Moving Vehicle"	"Real Time object detection and Tracking for Autonomous Driving Applications
Method	Deep convolutional neural network and Kalman filter for detection and tracking	YOLO for real-time object detection and tracking	Deep neural network for object detection and Kalman filter and Hungarian algorithm for object tracking
Dataset	Custom outdoor dataset with LiDAR and camera sensors	COCO dataset is used	KITTI dataset for object detection and tracking in autonomous driving applications
Result	Detection accuracy of 87.3% and tracking accuracy of 82.5% on outdoor dataset	Its processing speed is 45 frame per/sec	The system is appropriate for real-time applications because it is built to operate with low-latency input and output.

REFERENCES

[1] X.Zhang,Y Yang, and Y.Liu , "Infrastructure-Based Object Detection and Tracking for Cooperative Driving Automation: A Survey" 2022 IEEE Intelligent Vehicles Symposium (IV) ,IEEE

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection." arXiv preprint arXiv:1506.02640 (2015)."

[3] S.Wang,R.Clark and H.Wen, "Real Time object detection and Tracking for Autonomous Driving Applications"

[4] C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," in ICCV, 2015.

[5] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d objectdetection network for autonomous driving," in CVPR, 2017.

[6] Dundar, J. Jin, B. Martini, and E. Culurciello, "Embedded streaming deep neural networks accelerator with applications," IEEE Trans. Neural Netw. & Learning Syst., vol. 28, no. 7, pp. 1572–1583, 2017.

[7] R. J. Cintra, S. Duffner, C. Garcia, and A. Leite, "Low-complexity approximate convolutional neural networks," IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–12, 2018.

[8] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data." IEEE Trans. Neural Netw. & Learning Syst., vol. PP,no. 99, pp. 1–15, 2017.

[9] Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis." IEEE Trans.Neural Netw. & Learning Syst., vol. 23, no. 4, pp. 596–608, 2012.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation,"in CVPR, 2014.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only lookonce: Unified, real-time object detection," in CVPR, 2016.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in NIPS, 2015,pp. 91–99.

[13] D. G. Lowe, "Distinctive image features from scale-invariant key-points," Int. J. of Comput. Vision, vol. 60, no. 2, pp. 91–110, 2004.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR, 2016.

[15] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in ICIP, 2002.

[16] C. Cortes and V. Vapnik, "Support vector machine," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," J. of Comput. & Sys. Sci., vol. 13, no. 5, pp. 663–671, 1997.

[18] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, pp. 1627–1645, 2010.

[19] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zis-serman, "The pascal visual object classes challenge 2007 (voc 2007