Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

Advanced Intelligent Approaches for Robust Real Time Object Detection in Computer Vision

Mahesh D [1]

Research Scholar, Department of Computer Science Engineering,
Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P., India.

Dr. Harsh Lohiya [2]

Research Supervisor, Department of Computer Science Engineering,
Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P., India.

Keywords:

Object Detection, Background, Algorithms, Real-time, Accuracy.

ABSTRACT

Object detection is one of the most basic, popular, and difficult problems in computer vision. Researchers and computer scientists have put in a lot of time and energy over the last many decades trying to solve the object detection issue. In its most basic form, object detection is a tool for finding specific objects in moving images. Many studies have been conducted in this area since the turn of the century; nevertheless, current methods like as HOG, SIFT, SURF, etc., are not yet suitable for application in real-time detection because to their slowness and lack of precision. In addition, Convolution Neural Networks underwent a fast transformation during the deep learning period, which paved the way for a new avenue of research that has yielded several outstanding results to this day (e.g., YOLO, SSD, retina NET, etc.).

I. INTRODUCTION

Given that people can effortlessly identify a broad variety of items, regardless of their form, colour, texture, etc. To put it simply, object recognition is the study and practice of making computers execute tasks that humans do—namely, identify and recognise objects in images. Object recognition has found widespread use in many different fields. Things may be seen from various angles, at all sizes, and even while partly obscured from one's line of sight; this is common knowledge. Consequently, we need a method for object identification that remains constant regardless of the item's size, position, or orientation.

One area of computer vision that has persisted over the last half-century is object identification in pictures. The object detection method simplifies the process of identifying the kind and position of objects in visual media. With this innovation, it's possible to locate and identify each and every item in a video or picture.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

In the same way that a person can identify different items in a picture by comparing it to others, an object recognition system can do the same thing with an image. Variegated illuminations, partial occlusions, perspective shifts, appearance fluctuations, crowded backgrounds, and other picture issues continue to make object recognition a difficult process. Consequently, it is necessary to have an effective object recognition method that can both identify objects in these situations and solve these kinds of difficulties. A fast and accurate method for feature detection and extraction is a part of the object recognition technique.

There are essentially three phases to the object recognition method. As a preliminary step, we identify the image's corners and intersections as potential key-points. Step two involves using feature vectors, often called descriptors, to depict the areas around the critical locations. For effective identification, the descriptors need to be both unique and resistant to noise. The last stage is to match the descriptor vectors across all of the photos to identify the item.

II. METHODS OF OBJECT DETECTION

The various object detection methods are described as

Template Based Object Detection

Using the template picture, this approach can recognise the little sections of the image. Template matching is another name for this method. The mobile robot assesses the picture's quality by locating its edges and comparing them to the quality control image. In order to find the relationship between the template picture and the actual image, geometrical parameters are used. Multiple rounds of the geometrical parameters are used in the data picture for template matching. In the search pictures, the geometrical parameters are denoted as S(x,y), where (x, y) denote the coordinates of each pixel [18]. In order to put the strategies into action, search images are used to locate the templates. In order to calculate the quantity of products between the coefficients throughout the whole region covered by the template, its origin moves over every point of the search picture. All positions are evaluated based on their highest score. Filter mask is the template and spatial filtering is the procedure.

Part Based Object Detection

Data pertaining to an object's representation in its deformable state. The distorted arrangement is represented by the connections between each pair of pieces in the models, which are individually organised. For general recognition issues, these models are ideal as they ascertain how the qualitative descriptions seem visually.

Region Based Object Detection

Transforming an input picture into a directed graph according to a set of rules established by an algorithm. In the process of building the graph, the graph's properties are extracted, which represent the object's global shape information included in the input picture. This method improves the graph's processing speed by representing the traversal of the post-preserving graph. After running the algorithm through a dedicated database, we can see that it struggles with two issues: object class detection and retrieval of comparable images.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

Contour Based Object Detection

The picture database classifies the many things that may be captured by a single prototype image. Wherever the robots are, they may be detected by items passing through their viewfinders. Images are captured using cameras so that the items may be recognised and moved to their ultimate locations. This procedure is divided into two parts. In the first stage, we define the polynomial forms and pinpoint where the items are. The strong connections of the holes are present in almost all forms. Phase two involves prototypes, which identify item categories and determine relative orientation. An alternate strategy that relies on initial picture detection is combining the polygon technique with image segmentation. As a statistical method, the triangle approximation ensures accuracy.

Appearance Based Detection

Here, occlusion and clutter are handled by the object identification system that relies on 3D object recognition. Images and settings may make use of looks to enhance their visual appeal. The primary categories, which are based on two-dimensional perspectives of things, are local and global approaches.

Background Subtraction

This technique involves removing the foreground items from the picture frames while keeping the background organisations. Background subtraction in non-recursive approaches is accomplished using techniques such as frame differencing, median filtering, and linear predictive filtering. Using the backdrop model to estimate the video frames and storing statistical features of frames in a large amount of memory is the major goal of this technique. Background estimate based on input frames and median filter or Kalman filter approximation is maintained by the different approaches. You may think of the recursive technique as just iterating over the same items in a same fashion. When compared to non-recursive and computational methods, fresh video frames requiring less memory storage may be used to update the single backdrop model.

Foreground Detection Method

This method uses backdrop subtraction to isolate things from a representation of their surroundings. Foreground image recognition involves comparing the incoming video frames to the backdrop model. Following this, the picture pixels are extracted from the original picture frames. Therefore, the kind of approaches determines the rate of foreground detection. The alternative method involves detecting the foreground by using a typical statistics-based threshold. In addition, the problem with this approach is that it relies on the outdoors, which causes the geographical unpredictability of established thresholds. Shadow detection and background removal utilising density and Kernel estimates are used to compare the original background models.

III. OBJECT DETECTION ALGORITHMS

Traditional Algorithms in Object Detection

Object identification algorithms have always relied on manually created features since, before to 2012, there were insufficient sophisticated methods for representing pictures. In order to make up for the

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

disparity caused by the lack of available computing resources, researchers had to come up with vector representations of very complicated characteristics and use a variety of acceleration approaches.

- **SIFT**

Professor David G. Lowe of Canada first proposed this approach in 1999. Principal points that are fixed and do not change with size, rotation, or location may be found using SIFT. Additionally, it excels at recording illumination changes as well as transitions (such as resizing, rotating, cropping, and positioning). Consequently, this method is applicable to object.

- **HOG**

N. Dalal and B. Triggs first presented it in 2005. A dense grid of uniformly spaced cells is its primary architecture for doing computations. Although HOG has many potential applications, the main motivation for its development has been the identification of pedestrians. By repeatedly rescaling the input picture without altering the size of the detection window, the HOG detector can identify objects of varying sizes—its finest characteristic.

- **SURF**

Herbert Bay and others were the first to introduce it. The SURF method, developed in 2006, is a potent and quick tool for static local similarity representation and picture comparison. The SURF method is attractive for real-time applications like tracking and object identification because it quickly calculates operators using box filters. This method is far quicker than SIFT.

- **ORB**

Using ORB as a foundation, Li Xiaohong and colleagues presented a novel approach to object recognition in dynamic situations in 2012. In order to accomplish the ultimate motion objective, it employs the eight-parameter rotation model in conjunction with the least squares approach to determine the total motion parameters for motion compensation. The experimental results demonstrate that this approach improves detection levels and real-time efficiency while simultaneously incorporating SURF. Furthermore, it is able to consistently and swiftly follow objects in motion in real time.

Object Detection Algorithms Based on Deep Learning

While more conventional methods of discovery certainly have their uses, deep learning has revolutionised computer vision and is now ubiquitous in our lives. Deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have opened several doors to the development of new technologies, and neural-based approaches have improved accuracy and efficiency while adding new features.

Two primary types of deep learning-based object detection devices exist: single-stage sensor devices and two-stage sensor devices with two stages. In the two-stage sensor device, the first stage involves generating a limited collection of basic propositions. The second stage involves extracting feature vectors from these propositions and encoding them using deep convolutional neural networks. Finally,

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

predetermined class predictions are created. In contrast, a single-stage sensor device would often try to identify each region of interest as either a target item or a background, treating every point in the input picture as a possible target object. In comparison to two-stage sensor devices, which provide remarkable results on several publically accessible datasets, single-stage sensor devices are often much slower and less suitable to real-time object identification applications.

- **Two-Stage Detectors**

1) R-CNN

The acronym R-CNN refers to a Region Based Convolutional Neural Network, which was first introduced in 2014 by Girshick et al. Classification and localisation are the two main components of object detection. The bounding box area, also known as the "region of interest" or "ROI," is first identified by a controlled number of candidates found through a selective search. Afterwards, it utilises the area suggestions to locate items in the picture after individually extracting CNN features for classification from each region.

2) SPP-Net

His work and others' was first published in 2014. Bypass neural architecture SPP-Net eliminates the network's fixed size limitation by spatial hierarchy clustering. In particular, we build upon the previous convolutional layer by adding an SPP layer. After the SPP layer has collected features, it will generate an output of a set length and feed it into the fully linked layers. To rephrase, in order to prevent clipping or convolution at the beginning, we acquire some information at a deeper level of the network architecture, between fully connected layers and convolutional layers.

3) Fast R-CNN

Fast R-CNN is an object detector developed specifically in 2015 by Ross Gerchik, an AI researcher at Facebook and a former researcher at Microsoft. Fast R-CNN overcomes many problems with R-CNN. As the name suggests, one of the advantages of Fast R-CNN over R-CNN is its speed. In a Fast R-CNN architecture, as shown in Fig. 3, whole image is given as input along with object proposals. Several convolutional and max-pooling layers are used to process the image and produce conv feature map as output.

4) Faster R-CNN

S. Ren et al. presented Faster R-CNN technology in 2015, not long after Fast R-CNN was announced. An RPN (Region Proposal Network) fully convolutional layer speeds up R-CNN. This layer processes randomly sized pictures and may provide a set of suggestions at each feature map point. The generated feature map is used to generate feature vectors, which are then input into a classification layer. Then, a bounding box regression layer is used for item localisation, and lastly, object detection is achieved. As far as end-to-end, quasi-real-time deep learning detection models go, Faster R-CNN is among the first. Even though Faster R-CNN can detect more quickly than Fast R-CNN, computational inertia occurs at each level of the detection process.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

5) R-FCN

A fully convolutional network (R-FCN) that divides up the overall computing cost during the area classification stage was proposed by Dai et al. in 2016. Without using region-wise fully connected layers, R-FCN outperformed Faster RCNN. For one, R-FCN builds a location-aware point map to store the relative positions of various categories; for another, it employs a location-aware clustering layer to extract characteristics from the spatially aware area by storing the relative positions of each target region.

6) FPN

To facilitate object recognition at various levels, Lin et al. (2017) presented an FPN method that combines characteristics from deep and shallow layers. In order to build strong semantic features from deeper layers, it is essential to construct features that are spatially correct. Improved detection performance is achieved by including this FPN into the conventional Faster R-CNN model. It is worth mentioning that FPN has become an essential component of the most recent.

7) Mask R-CNN

Regarding picture and sample segmentation, Mask R-CNN is a top-tier Convolutional Neural Network (CNN). Its 2017 introduction was by He et al. Performing pixel-level segmentation is the goal of the Mask R-CNN, which stands for Mask Regional Convolutional Neural Network. The difference between Faster R-CNN and Mask R-CNN is that the latter adds a step called Mask. To find out whether a pixel is an object component, it analyses each one individually.

- **Single Stage Detectors**

1) OverFeat

Prior to this, Sermanet et al. As one of the first methods to use deep learning for object identification, it outperforms R-CNN in speed but falls short in accuracy. In the last pooling layer of a deep convolutional neural network (DCNN), OverFeat employs a rapid multi-scale sliding window to extract points. Assigns stains to scores and estimates the rating for each stain.

2) YOLO

As early as 2015, Redmon et al. Dividing the picture into several grid cells is the fundamental principle of YOLO technology. After that, every single cell in the retina is subjected to algorithms for localisation and classification. It is the job of the object hub to determine which class label or label is associated with each network item. Because the whole picture is trained at once, detection efficiency is greatly enhanced. Because it does not need complicated pipelines, it is quick.

3) SSD

As seen in Chart 6, the Single-Shot Detector (SSD) variant was released in 2015. When it comes to object detection, SSD is a one-stage paradigm with only one pipeline. In 2016, Wei Liu et al. were the first to present SSD technology, which was developed to address the issues of R-CNN's low detection

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

rate and YOLO's poor accuracy. Hierarchical feature extraction is the foundation of SSD network architecture. Since SSD consolidates all computations onto a single grid, it simplifies approaches that include object recommendations by doing away with suggestion creation, final pixels, or resampling. Because of this, the SSD may be readily integrated into systems that need an object detection component and trained with ease. SSD outperforms other single-stage algorithms in accuracy while requiring a lower input picture size.

4) RetinaNet

Lin et al. introduced the RetinaNet paradigm for single object identification in 2018. When training with an imbalanced set of classes, the attention loss function is used to restore parity. Concentration loss modifies the entropy loss of concentration in hard negative instances. A central network and two specialised subnets for individual tasks make up RetinaNet. An autonomous convolutional network, the backbone computes the convolutional feature map across the whole input picture. As a result of the backbone's output, the first subnet sorts convolutional objects; the second subnet applies bounding box convolutional regression. For single-stage dense sensing, the authors advise using the two subnets' straightforward architecture.

IV. CONCLUSION

Current research and technology are making tremendous strides in several domains, one of which is object identification, an essential capability of computer vision systems. Object identification and deep learning (neural networks) have both seen a great deal of outstanding research in recent years, according to this paper's extensive literature survey. technology is advancing at a rapid pace, but there are still some challenges to robust object detection. Many of these challenges have been partially or completely solved, but this paper's findings suggest that there is still room for improvement. Some of these challenges include: occlusion, where objects are overlapped or unclear and difficult to identify correctly; viewpoint variation, where objects appear completely different from different angles; efficiency, speed, computational power, and deformation object detection, where objects (such as the human body) can change their shape and be difficult to detect accurately. Consequently, there is still a lot of space for development in this domain.

REFERENCES

[1] D. Diwakar and D. Raj, “Recent object detection techniques: A survey,” *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 2, pp. 47–60, 2022.

[2] Y. Xiao, Z. Tian, J. Yu, Y. Zhang, S. Liu, S. Du, and X. Lan, “A review of object detection based on deep learning,” *Multimedia Tools and Applications*, vol. 79, no. 33, pp. 23729–23791, 2020.

[3] F. Xiao, W. Deng, L. Peng, C. Cao, K. Hu, and X. Gao, “Multiscale deep neural network for salient object detection,” *IET Image Processing*, vol. 12, no. 11, pp. 2036–2041, Nov. 2018.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

[4] D. Patel and P. K. Gautam, "A review paper on object detection for improve the classification accuracy and robustness using different techniques," *International Journal of Computer Applications*, vol. 112, no. 11, pp. 975–8887, 2015.

[5] S. Kamate and N. Yilmazer, "Application of object detection and tracking techniques for unmanned aerial vehicles," *Procedia Computer Science*, vol. 61, no. 3, pp. 436–441, 2015.

[6] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, "A survey on object detection and tracking methods," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 2, pp. 2970–2979, 2014.

[7] S. Awan, *Object Class Recognition Using Global Shape Descriptors in 3D*, Ph.D. dissertation, vol. 2, no. 1, pp. 345–389, 2014.

[8] X. Li, C. Xie, Y. Jia, and G. Zhang, "Rapid moving object detection algorithm based on ORB features," *Journal of Electronic Measurement and Instrument*, vol. 27, no. 5, pp. 455–460, 2013.

[9] R. Oji, "An automatic algorithm for object recognition and detection based on ASIFT keypoints," *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 3, no. 5, pp. 29–39, Oct. 2012.

[10] M. Z. Kurian and C. M. MV, "Various object recognition techniques for computer vision," *Journal of Analysis and Computation*, vol. 7, no. 1, pp. 39–47, 2011.

[11] K. Schindler and D. Suter, "Object detection by global contour shape," *Pattern Recognition*, vol. 41, no. 12, pp. 3736–3748, 2008.