iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58810

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

A Study of YOLO-Based Object Detection for Visually Impaired Individuals

Anita[1], Rupali Chandrakar[2], Dr. Shikha Pandey[3]
[1]*Research Scholar, Department Of Computer Science & Engineering, RSR Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India*
[2, 3]*Assistant Professor, Department Of Computer Science & Engineering, RSR Rungta College Of Engineering & Technology, Bhilai, Chhattisgarh, India*

Abstract: ***An essential area of investigation involves recognizing and tracking objects, given the frequent changes in object motion, scene dimensions, occlusions, appearance, ego-motion, and lighting variations. Effective object tracking heavily relies on feature selection due to its critical role. This process is integral to numerous real-time applications like vehicle detection and video surveillance. To address detection challenges, tracking methods need to be adept at handling object movement and appearance changes. Among these methods, tracking algorithms play a pivotal role in smoothing video sequences and receive significant attention. However, only a few approaches leverage previously gathered data on object attributes such as shape and color. This study delves into a tracking algorithm that encompasses all these object properties. Its aim is to investigate and assess past methodologies for tracking and detecting moving objects across various stages of video sequences. Additionally, it aims to identify existing challenges and propose innovative strategies to enhance object tracking throughout video frames.***
Keywords: Object tracking, object recognition, statistical analysis, object detection, back ground subtraction, performance analysis, and optical flow.

I. INTRODUCTION

In recent years, there has been significant progress in miniaturization, coupled with a decrease in camera costs, leading to a preference for building extensive networks of cameras. This proliferation of cameras has facilitated innovative signal processing applications utilizing multiple sensors across wide areas. Object tracking, a relatively novel technique, involves leveraging cameras in various video sequences to locate moving objects in real-time (Kothiya and Mistree, 2015). The primary goal is to establish a connection between target objects, their attributes, and their positions in subsequent video frames. Consequently, object categorization and detection are crucial aspects of object tracking in computer vision applications. Tracking serves as the initial step in identifying or detecting moving objects within frames. Subsequently, the detected objects may be classified as moving trees, birds, people, vehicles, etc. However, object tracking using video sequences poses significant challenges in image processing, including occlusions, complex object motions, real-time processing requirements, and inappropriate object shapes.

Despite these challenges, object tracking offers numerous advantages, including traffic monitoring, robot vision, surveillance, video communication, public space monitoring (e.g., subway stations, airports), and event tracking (Kim, 2007; Lowe, 2004; Ojha and Sakhare, 2015; Yilmaz et al., 2006). Consequently, achieving the best balance of computational power, network bandwidth, and system accuracy is essential for specific applications. The level of cooperation among cameras for data collection, distribution, and processing determines the computational and communication costs. This cooperation is crucial for validating decisions and minimizing estimation errors and ambiguities.

In order for a machine to effectively interact with humans, it is imperative that it possesses the ability to identify suspicious objects and comprehend their actions within a given context. The current approach to analyzing and identifying such objects often relies on specific indicators associated with the objects in question, which limits the widespread adoption of this technology. The aim of this study is to explore and evaluate the existing method of object tracking using video sequences that encompass various stages.

The subsequent discussion outlines three crucial phases in video analysis:

1) Identification of the targeted object within a moving sequence.
2) Tracking the object from one frame to another.
3) Continuous tracking of the object across multiple cameras.

©IJRASET: All Rights are Reserved

II. THE STATE OF THE ART

Object detection, a prominent computer technology closely related to computer vision and image processing, is dedicated to identifying objects or instances belonging to specific classes (such as humans, flowers, or animals) within digital images and videos. This technological concept was initially introduced by Bhumika Gupta et al. (2017). Object detection has been extensively explored across various applications, including character recognition, vehicle detection, and facial recognition. It serves diverse purposes such as surveillance and retrieval. The aim of this investigation is to enhance the efficiency and precision of object detection by introducing several fundamental principles employed in this field. These principles are demonstrated using the OpenCV library in Python 2.**7.**

Kartik Umesh Sharma and colleagues (2017) introduced a system for object identification that identifies real-world objects present in digital images or videos. These objects may belong to various classes such as people, vehicles, or other categories. A complete system for object detection includes components like a model database, feature detector, hypothesis generator, and hypothesis verifier. These components are essential for the system to effectively identify objects. This article provides an overview of numerous strategies employed in object recognition, including object localization, categorization, feature extraction, and appearance analysis from images and videos. These techniques are thoroughly discussed throughout the study, drawing insights from literature analysis and highlighting significant concerns in object detection. Additionally, the article offers resources such as source codes and accessible datasets to assist new researchers in the object detection field. It also proposes a potential method for detecting multiple classes of objects.

Mukesh Tiwari and colleagues (2017) presented object detection and tracking as a critical area of research due to the dynamic nature of object motion and variations in scene conditions. Feature selection is emphasized as a crucial aspect of object tracking, relevant to various real-time applications such as video surveillance and vehicle perception. Tracking based on object movement and appearance addresses detection challenges. The study focuses on discussing and analyzing a tracking algorithm that incorporates multiple object parameters. Its objective is to investigate and evaluate methods used in the past to track and detect moving objects using video sequences across different stages. Furthermore, the study aims to identify gaps and propose new strategies to enhance object tracking throughout video frames.

According to Aishwarya Sarkale and colleagues (2018), humans possess remarkable visual discrimination abilities, but object recognition poses challenges for machines. This led to the development of neural networks, computational models inspired by the brain, to aid in object detection and identification. The study discusses various types of neural networks such as Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Faster R-CNN, 3D Convolutional Neural Networks (3D-CNN), Recurrent Neural Networks (RNN), among others, along with their respective accuracies. Analysis of multiple research publications reveals varying degrees of accuracy for different neural network architectures. Upon analyzing and comparing Neural Networks, it can be concluded that Artificial Neural Networks (ANN) demonstrate the highest level of accuracy for object detection in the provided test scenarios.

Karanbir Chahal and colleagues (2018) defined "object detection" as the process of identifying, localizing, and categorizing objects within images, which is crucial for vision-based software systems and offers a wide range of potential applications. The aim of their research is to conduct a comprehensive analysis of contemporary object detection methods utilizing deep learning techniques. This study delves into various aspects including training approaches, performance metrics, trade-offs between speed and model size, and different algorithms. It particularly focuses on two categories of object detection algorithms: single-step detectors like SSD and two-step detectors like Faster R-CNN. Additionally, the research explores the development of new lightweight convolutional base architectures to address the challenge of creating detectors that are both portable and efficient on low-power devices. Ultimately, through a thorough analysis of the strengths and weaknesses of each detector, the study presents the current state-of-the-art in object detection.

Richard Socher and colleagues (2018) proposed that advancements in 3D sensing technology enable the easy capture of both color and depth images, which, when combined, can enhance object detection. Most existing techniques rely on components specifically designed for this new 3D modality. They introduce a model for learning features and categorizing RGB-D images based on a combination of convolutional and recursive neural networks (CNNs and RNNs). The CNN layer extracts low-level features that are translationally invariant, serving as inputs for multiple RNNs structured as fixed trees to generate higher-order features. RNNs integrate convolution and pooling operations into a single hierarchical and efficient process. Notably, they found that RNNs can construct useful features even with completely arbitrary weights. Their model achieves state-of-the-art performance on a standard RGB-D object dataset, being more accurate and faster during training and testing compared to similar architectures such as two-layer CNNs, surpassing other available models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Yordanka Karayaneva and colleagues (2018) highlighted the use of robots as social peers in schools worldwide to engage with children and young students, enhancing their overall learning experience. This approach has been shown to significantly improve academic performance.

They utilized the humanoid robot NAO, capable of recognizing objects, colors, shapes, typed text, handwritten numerals, and operators.

The robot's ability to recognize written words facilitates matching gestures in sign language, while five different classifiers, including neural networks, are employed for handwriting recognition. Evaluation using robot-captured photographs, including sign language motions, revealed object identification algorithm accuracy ranging from 82% to 92%. The accuracy of the handwriting recognition classifiers ranged from 87% to 98%. This project presents a promising option for providing emotional support to younger students and children.

Abdul Muhsin M and colleagues (2019) advocated for the independence of every individual, particularly emphasizing the importance for disabled individuals. In recent years, technology has increasingly focused on aiding disabled individuals to gain greater control over their lives. Their work introduces an assistive system for the visually impaired, aiming to provide awareness of surroundings by employing YOLO for fast object recognition in photos and video streams using deep neural networks, implemented with OpenCV in Python on a Raspberry Pi3. The results indicated that the proposed model successfully achieved its objective of enabling blind users to navigate indoor and outdoor environments by offering a user-friendly device based on person and object recognition models.

Geethapriya. S and colleagues (2019) proposed the objective of detecting items using the You Only Look Once (YOLO) technique, highlighting its advantages over alternative object identification methods. Unlike other algorithms such as Convolutional Neural Networks and Fast Convolutional Neural Networks, YOLO examines the entire image rather than parts, accomplishing this much faster than previous methods by predicting bounding boxes and class probabilities using a neural network.

R Sujeetha (2020) suggested that if implemented, the proposed object recognition and tracking could become a significant and dynamic area in computer vision. Researchers have developed simplified and efficient solutions for government surveillance, security tracking, and various other applications. However, real-time object identification and tracking implementations face challenges due to the need for optimized performance and efficient time component detection. Tracking multiple objects further complicates the task.

Various methodologies have been devised, but there is room for improvement. This approach utilizes TensorFlow and OpenCV libraries along with the CNN algorithm to assign labels to identified layers and verify their accuracy. Additionally, it enables real-time detection and simulation of these layers using external hardware, ultimately providing highly optimized and effective algorithms for object tracking and detection.

Mahesh Pawaskar and colleagues (2023) highlight the numerous challenges faced by visually impaired individuals when interacting with their immediate surroundings. Their aim is to propose a device that assists visually impaired individuals in both navigation and obstacle perception. They propose an operational model incorporating an ultrasonic sensor and a microcontroller device integrated into a walking stick. The detection and monitoring algorithms are designed to leverage image and motion data for protection and surveillance applications.

Well-known algorithms for object detection include You Only Look Once (YOLO), region-based Convolutional Neural Network. (RCNN), and Faster RCNN (F-RCNN). While RCNN boasts superior accuracy compared to other algorithms, YOLO outperforms it in terms of speed, albeit at the cost of some accuracy.

III. PROPOSED METHOD: OBJECT DETECTION AND TRACKING

The proposed method for object detection and tracking aims to develop a robust and efficient system capable of accurately identifying and monitoring objects in real-time scenarios. The method integrates various techniques and technologies to achieve optimal performance in object detection and tracking tasks.

A. OpenCV

The proposed method for object detection and tracking involves utilizing OpenCV, a comprehensive library covering image recognition, deep learning, and image analysis. OpenCV has the capability to detect objects, people, and even handwriting from both still images and video footage.

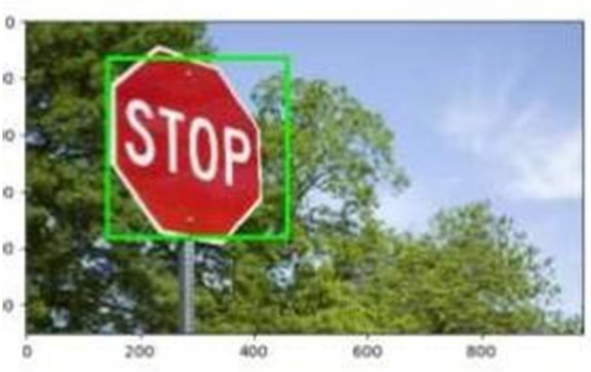


Fig.1 -Image Recognition Using OpenCV

Object detection involves training a cascade function, which begins with labeling a large number of images as positive or negative examples.

Once the classifier is trained, distinguishing characteristics known as "HAAR Features" are extracted from the training images. These features consist of rectangular shapes with variations of bright and dark pixels. The classifier is trained to minimize error rates by selecting relevant features associated with the object of interest. This approach allows for efficient object detection by focusing on pertinent image characteristics while disregarding irrelevant features.

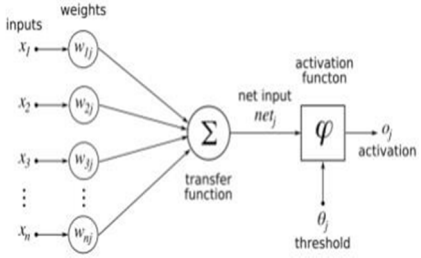


Fig.2: Convolution Neural Networks (CNN)

The context take supalarge portion of the image; The item to be viewed is only a small portion of the picture. Cascaded classifiers are used to speed up the detection process. If even a single negative feature is detected in a region of an image during this step, the algorithm goes onto the next region after ignoring the region for further processing. The requisite object in the image is the only area that contains all of the identifying features. The requisite object in the image is the only area that contains all of the identifying features.

B. CNN Architecture

In a convolution neural network, there is a layer for information processing as well as an output layer, in addition to a number of hidden layers. Typically, a CNN's hidden layers are made up with a sequence of convolution layers that start with an increase or another dot product. The activation function is typically a RELU layer, and it is consequently followed by additional convolutions such as pooling layers, completely associated layers, and normalization layers. These layers are referred to as hidden layers due to the fact that their sources of input and output are masked by the activation function and the last convolution layer in the network..

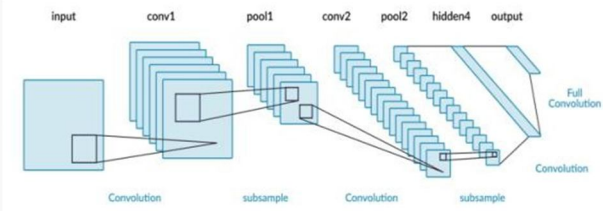


Fig.3-The Basic Component of Artificial Neural Networks, The Artificial Neuron, has the Following Structure.

C. SVM Classifier

Support A form of supervised machine learning method known as a vector machine provides knowledge analysis that may be used for classification and multivariate analysis. SVM is often utilised for classification purposes, however it will also be utilised for regression. The graphing that we do within the n-dimensional space is highlighted below. In addition, the value of each individual feature is equivalent to the value of the precise coordinate. Then, we choose the ideal hyper-plane that clearly demarcates the difference between the two classes. It's not hard to understand how support vector machines (SVMs) accomplish their task because to their obvious underlying idea. To begin, let's take a look at a sample problem once you've developed a hyper plane that sorts the data set into categories. Imagine that you have to differentiate between red triangles and blue circles in a certain data collection. Your task is to do so. Your objective is to draw a line that divides the information into two categories, establishing a division between the blue circles and the red triangles..

D. Single Shot Detector (SSD) Algorithm

SSD is a popular object detection algorithm that was developed in Google Inc. [1]. It is based onthe VGG-16 architecture. Hence SSD is simple an deasier to implement.

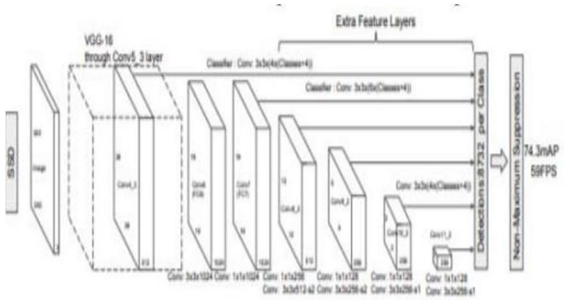


Fig.4. VGG-16 SSDModel.

The VGG 16 SSD model may be seen in Figure 4. Convolutional processing is applied to a group of default boxes so that they can move across many feature maps. During the process of prediction, a score is created only in the event that an item found matches the criteria of one of the object classifiers. The shape of the item is altered so that it conforms to the localization box. There is a prediction made on the confidence level and shape offsets for each box. During training, the default boxes and the ground truth boxes are compared and matched. SSD architecture discards the layers that are completely linked to one another. To calculate the model loss, a weighted sum of the confidence loss and the localization loss is added together. The localization loss is a measurement of how far the box that was anticipated deviates from the box that really existed. Confidence is a measurement of the degree to which a system is confident that an anticipated outcome will occur. Object refers to the real thing being discussed. It is much easier to train with MobileNets because to SSD's elimination of feature resampling and encapsulation of all processing within a single network. SSD is a method that does explicit region suggestions and pooling. It is quicker than YOLO, which is the other way (including Faster R-CNN).

E. Mobile Nets Algorithm

Mobile Nets makes use of depth-wise separable convolutions, which is a technique that assists in the construction of deep neural networks. The Mobile Nets architecture is most suited for use in situations where there is little need for process control, as these situations typically include portable and embedded vision-based systems.

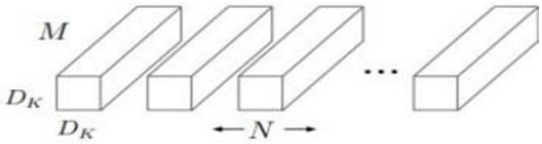


Fig.5.Normal Convolution[2]

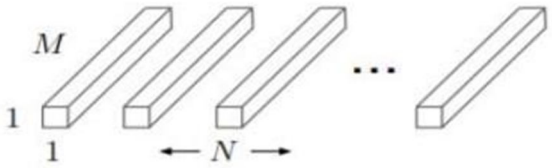


Fig.6.DepthwiseConvolutionFilters[2]

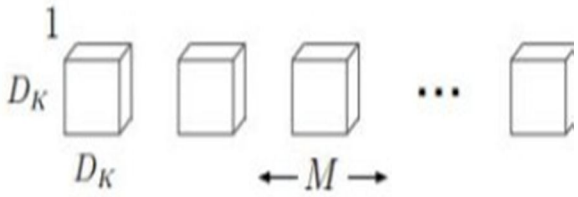


Fig.7.1×1Convolutional Filters[2]

The called Point wise primary goal of Mobile Nets is to minimize the occurrence of delay while simultaneously developing tiny neural networks as a parallel process. Its exclusive focus is on size, while speed is given relatively little consideration. Mobile Nets are built using depth-wise separable convolutions as its building blocks. After the standard convolution, the input feature map is divided up into a number of different feature maps [2].

Convolution in the context of Depth wise Separable Convolution [2].

When compared to the work done by the network using normal convolutions having the same depth, this model's implementation of depth-wise separable convolutions results in a sizeable reduction in the total number of parameters. This is achieved through the model's utilization of depth-wise separable convolutions. The decrease in parameters causes the construction of a lightweight neural network, which can be seen in figures 5 through 7 of the figure.

IV. CONCLUSION

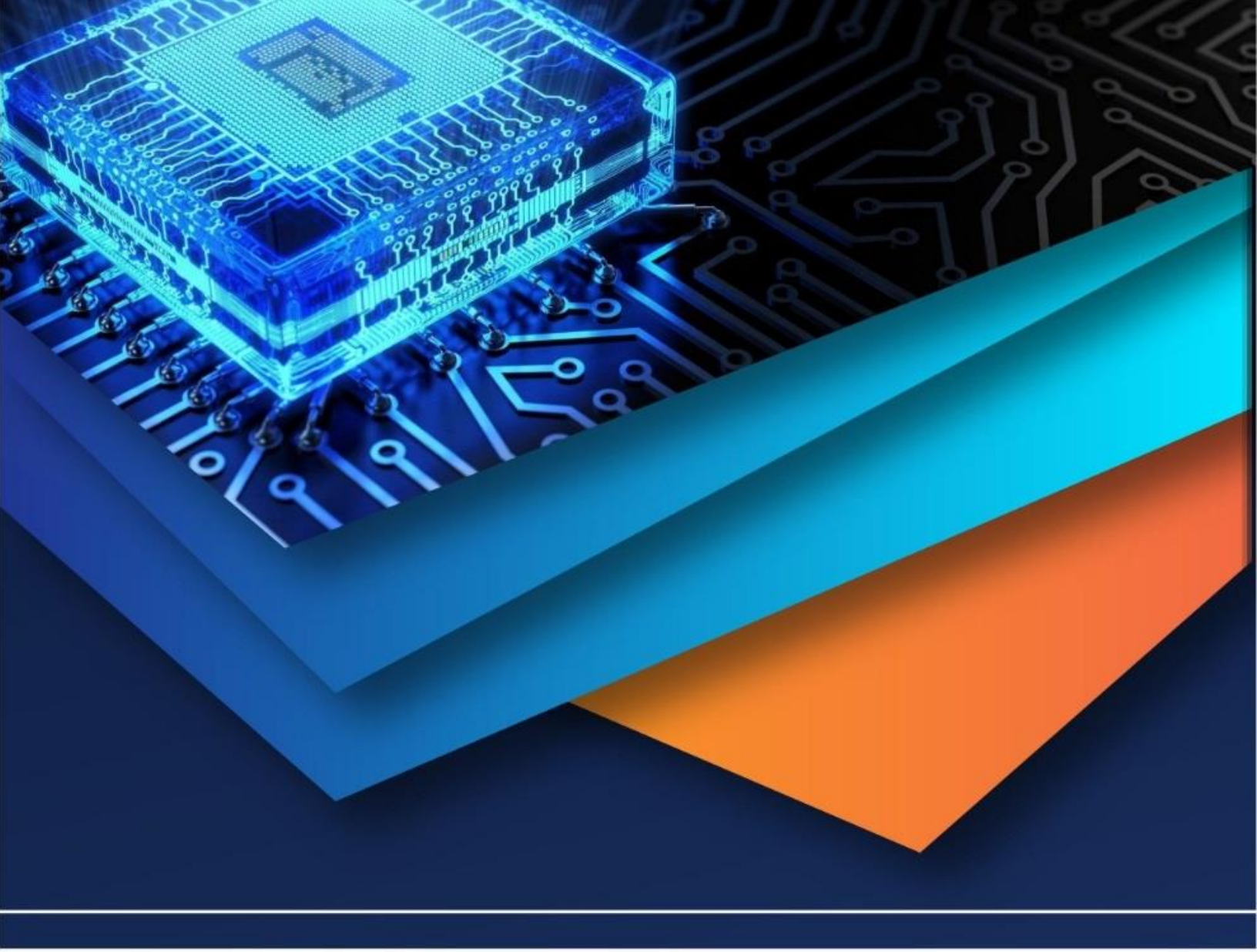
This study presents a comprehensive review of diverse algorithms for object detection, tracking, and identification, alongside feature descriptors and segmentation methods that operate on video frames. These methodologies aim to enhance object recognition through innovative concepts. Additionally, the study delves into the tracing of objects across video frames, providing theoretical justification for this process. The bibliography's content stands out as the primary contribution to the research, offering insights that could spark the emergence of a new field of study. We have examined the limitations of various methodologies and discussed their potential future directions. Furthermore, we have identified techniques that yield accurate results but entail high computational complexity, notably statistical approaches, background removal, and temporal differencing coupled with optical flow. However, addressing challenges such as abrupt changes in lighting, intense shadows, and object occlusions requires a heightened focus on managing these aspects within the proposed approach.

REFERENCES

[1] Abdul Muhsin M, Farah F. Alkhalid, Bashra Kadhim Oleiwi, “Online Blind Assistive System using Object Recognition”, International Research Journal of Innovations in Engineering and Technology (IRJIET), Volume3, Issue12, pp 47-51, December–2019.
[2] Aishwarya Sarkale, Kaiwant Shah, Anandji Chaudhary, Tatwadarshi P.N., “A Literature Survey: Neural Networks for Object Detection”, VIVA-Tech International Journal for Research and Innovation Volume1, Issue1 (2018) ISSN(Online): 2581-7280ArticleNo.9.
[3] Bhumika Gupta, Ashish Chaube, Ashish Negi, Umang Goel, “Study on Object Detection using OpenCV Python”, International Journal of Computer Applications Foundation of Computer Science (FCS), NY, USA, Volume 162,Number 8, 2017
[4] Geethapriya.S, N.Duraimurugan, S.P.Chokkalingam, “Real Time Object Detection with Yolo”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249–8958,Volume-8,Issue-3S, February2019.
[5] Karanbir Chahal1 and KuntalDey,“A Survey of Modern Object Detection Literature using Deep Learning”, International Research Journal of Engineering and Technology (IRJET),Volume8, Issue 9, 2018.
[6] Kartik Umesh Sharma and Nilesh singh Thakur, “A Review and an Approach for Object Detection in Images”, International Journal of Computational Vision and Robotics, Volume 7, Number1/2, 2017.
[7] Mukesh Tiwari, Dr. Rakesh Singhai, “A Review of Detection and Tracking of Object from Image and Video Sequences”, International Journal of Computational Intelligence Research,Volume13, Number5 (2017).

[8] R. Sujeetha, VaibhavMishra, "Object Detection and Tracking using Tensor Flow", International Journal of Recent Technology and Engineering (IJRTE)ISSN:2277-3878, Volume-8, Issue-1,May2020.
[9] Richard Socher, BrodyHuval, Bharath Bhat, Christopher D. Manning, AndrewY.Ng, "Convolutional Recursive Deep Learning for 3D Object Classification", International Conference on Computational Intelligence and Communication Networks,2018.
[10] Yordanka Karayaneva and Diana Hintea, "Object Recognitionin Python and MNIST Dataset Modificationand Recognition with Five Machine Learning Classifiers", Journal of Image and Graphics, Volume6, Number1, June 2018.
[11] Mahesh Pawaskar, Sahil Talathi, Shraddha Shinde, Digvijay Singh Deora "YoloV4 Based Object Detection for Blind Stick"International Journal of Innovative Science and Research TechnologyVolume 8, Issue 5, May – 2023.
[12] Mais R. Kadhim, Bushra K. Oleiwi "Blind Assistive System Based on Real Time Object Recognition using Machine Learning"Engineering and Technology Journal 40 (01) (2022) 159-165.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)