iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58510

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 II Feb 2024- Available at www.ijraset.com

Object Detection Using Image Processing For Blind Persons: A Review

Aditya Gandhi[1], Ayush[2], Devansh Srivastava[3], Ambuj Tripathi[4]
ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Abstract: The fundamental idea of detection of objects for videos is confirming an object's existence in a series of images and maybe pinpointing it specifically for identification. Monitoring an item's existence, location, size, form, and other physical and Changes in time throughout a video sequence is known as object tracking. The challenge of reproducing the designated area in consecutive frames of a set of photographs captured at closely spaced time intervals is referred to as the temporal correspondence problem. and it must be solved in order to do this. The two processes in question are interrelated. The foundation of tracking is detection, which often begins with the detection of objects. To aid and validate tracking, it is frequently required to identify an item repeatedly in a following image sequence. In light of this, the present review describes how object recognition using image processing might assist the blind.
Keywords: Object Detection, Image Processing

I. INTRODUCTION

A fundamental area of study in computer vision, deep learning, artificial intelligence, comes under object detection. Advanced computer vision tasks, such tracking of targets, detection of events, analysis of behaviour, and scene semantic comprehension, require it as an essential prerequisites. It seeks to precisely determine the category, locate the area of interest inside the image, and provide bounding box for every target. It is used in intelligent videography surveillance video ,also in image retrieval and automated driving [1], analysis of medical images [2], examination of industrial field.[3] Preliminary processing, window sliding, character extraction, choosing characteristics, featured classification, and post processing are the six basic processes of conventional techniques for manually extracting features. These steps are often applied for certain identification purpose. The primary drawbacks include limited data size, low portability, insufficient relevance, high temporal complexity, redundancy in windows, vulnerability to diverse changes, and performance limitations under specific conditions.

In the year of 2012, The AlexNet image classification model, utilizing convolutional neural network (CNN) architecture, was introduced by Krizhevsky [4] and colleagues. It emerged prominently in image classification competitions involving large datasets ImageNet [5], With an enormous accuracy using conventional methods, they emerged victorious in the competition. Numerous academics have started by the use of deeper convolutional neural networks for the subject identification problems and also put out a number of top-notch techniques. The two main types of detection algorithms are the two-level detection algorithm and process constructed on regression & also single-level observation strategy structured on region proposal.

In 2014, the R-CNN [6] Girshick presented a method, the first genuine convolutional neural network-based target detection model. An enhanced RCNN model attains a 66% map. For each image that has to be recognized, the model initially extracts about 2000 area recommendations using Selective Search. Subsequently, the retrieved proposals' sizes undergo uniform scaling to yield a fixed-length feature vector. These features are then fed into the SVM classifier to perform classification. Ultimately, the regression function of the bounding box is carried out by training a linear regression model. Although the R-CNN's accuracy is much higher than that of the conventional detection approach, its computation efficiency is too poor and its computation volume is too enormous.

Introduced in 2015, Spatial Pyramid Pooling (SPP) addressed efficiency and fixed input size issues in RCNN. It conducts feature extraction once per image, followed by a spatial pyramid pooling layer for feature vector extraction. However, it inherits RCNN's complexity, involving multiple stages and training regressors and SVM classifiers.

The Faster R-CNN [9] Ren's approach replaces the earlier Selective Search technique for generating region proposals with region proposal networks. The model consists of two modules: the Fast R-CNN detection method and a fully convolutional neural network module that generates all region proposals.

These two modules share a collection of convolutional layers. The last Shared the feature map for the RPN network's input is created. Faster R-CNN has convolutional layer receives the input picture after it has passed through the outstanding detection accuracy, but it still cannot achieve real-time detection.

CNN network. In order to create a higher dimensional feature map, the picture is first transmitted forward to the particular convolutional layer, after which frames per second on VOC2007. However, SSD performs poorly for small targets in terms of classification, and because features mappings across various scales are independent, boxes of different sizes might occasionally recognize the same object at the same time.

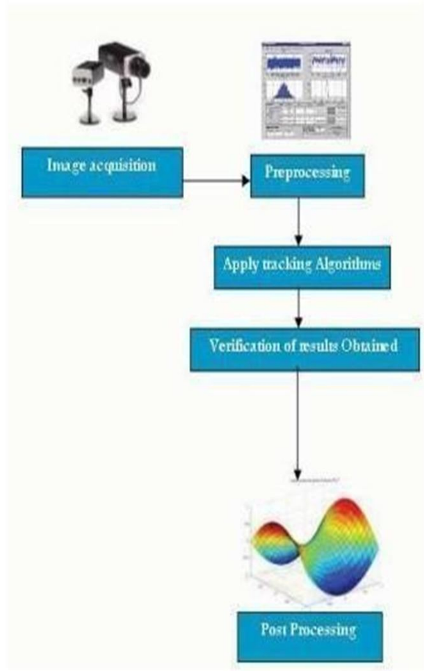
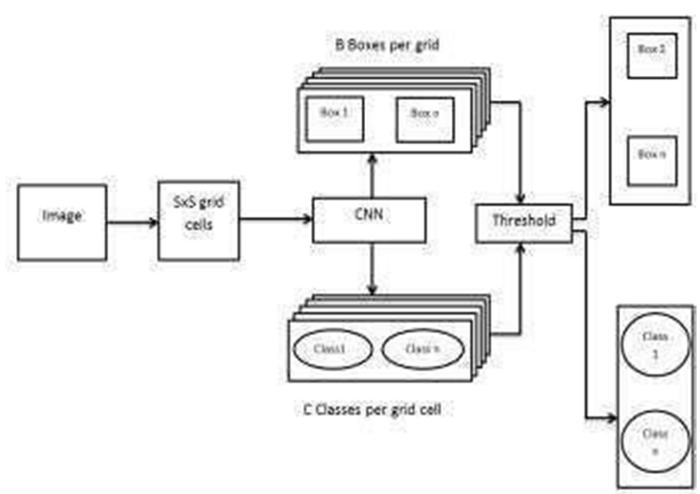


Figure 1-1 Block Diagram of system

Alexey Bochkovskiy came up with the concept for the YOLOv4 [14] in 2020, which breaks the previous record with the perfect balance of speed and accuracy. In theory, YOLOv4 is not all that new. It integrates Weighted Residual Connection, Cross Stage Incomplete connection, Cross small Batch Normalization, Selfadversarial training, Mish activation, Mosaic data enhancement, Drop Block, and CIou. It is based on the original YOLO detection framework. In order to expand the receptive area and identify the most important context factors, an SPP module was implemented based on the decision to use CSP Darknet53 as the backbone network. Meanwhile, YOLOv4 uses PANet as the path aggregation technique instead of FPN, as utilized in YOLOv3, and maintains the head structure of YOLOv3. The accuracy of the YOLOv4 is 10% and 20% higher.



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 II Feb 2024- Available at www.ijraset.com

II. ONE-STAGE TARGET DETECTION ALGORITHM

YOLOv1, Joseph Redmon presented the YOLOv1 [10] object detection model in 2016. The extraction method of region proposal is not necessary for the YOLOv1 detection model. All of the detection model is nothing more complex than a CNN network structure. Its main concept is to feed the network the whole graph as input and have the output layer deliver the bounding box's category and position right away. An S*S grid is first created from the picture, then for each grid cell, a B bounding box and confidence scores are predicted. That is, the entire B*(4+1) values are predicted by each cell. Its complete real-time detection speed on a single

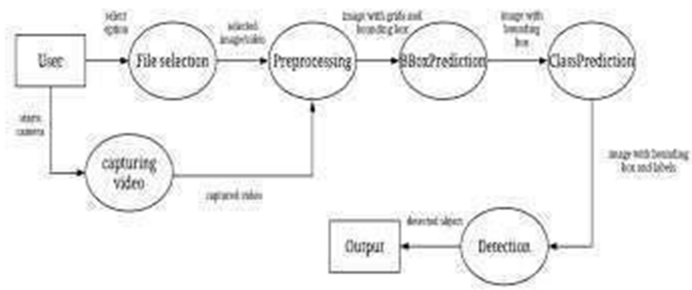


Figure 2: YOLO Architecture

TitanX is 45 frames per second. Yolo, on the other hand, generates less background mistakes but performs poorly in recognition when. The object identification model YOLOv3 [12], put out by Redmon, is by far the best balanced in terms of both detection accuracy and speed. Regarding category prediction, the primary function of YOLOv3 is to convert the initial single-label classification into a multi-label classification. A logistic regression layer is substituted for the original softmax layer, which was utilized for singlelabel multi-classification. Simultaneously, the model combines several scales to make predictions.. It uses a technique called up sampling fusion, which is similar to FPN, and fuses three scales in the end, greatly enhancing the detection of tiny objects. This model's network structure uses the Darknet53 deeper feature extraction network. The precision of detection has not increased considerably, especially when IOU>0.5, despite the fact that the YOLOv3 model significantly increases the speed of detection and the detection effect of tiny targets.

III. OBJECT DETECTION USING IMAGE PROCESSING

The aircraft commonly referred to as Unmanned Aerial Vehicles (UAVs) operate autonomously, without the need for a pilot on board. These vehicles can be controlled remotely from a ground control station or can navigate independently using advanced onboard systems or preprogrammed flight plans. UAVs are utilized for various purposes, including reconnaissance and assault missions. To distinguish them from missiles, UAVs are defined as aircraft capable of sustained flight at regulated altitudes, powered by either jet or reciprocating engines. Unlike cruise missiles, which are considered weapons themselves, UAVs serve as platforms for various missions. In some cases, the term UAV is expanded to UAVS (Unmanned Aircraft Vehicle System). The Federal Aviation Administration (FAA) introduced the term Unmanned Aircraft System (UAS) to emphasize that these systems consist of more than just the aircraft itself, including ground stations and other components. However, despite this official terminology, UAV remains widely used in everyday language due to its familiarity. As UAV capabilities continue to advance, countries continue to invest in research and development, leading to further advancements and expanding their range of tasks beyond intelligence, surveillance, and reconnaissance (ISR). These tasks now include combat search and rescue (CSAR), communication relay, electronic warfare (EW), strike missions, and suppression/destruction of enemy air defenses (SEAD/DEAD), among others.

A. Computer Vision And Object Detection

Images and videos are ubiquitous online, particularly on social media and photo-sharing platforms. Vision research has been revolutionized by machine learning and statistics, facilitating the identification, categorization, and tracking of objects and events in visual media to understand real-world scenarios.

Computer vision, powered by algorithms and machine learning, drives applications such as image search, photo management, medical image analysis, and robot navigation. In computer vision, scenes are perceived as compositions of objects against a background, with object recognition and detection being crucial tasks.

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 II Feb 2024- Available at www.ijraset.com

Object detection identifies an object's position and presence within an image, while object recognition assigns it to a specific class based on training data. Object detection methods vary between soft and hard detection, with the former identifying objects simply and the latter precisely locating them in position and presence. Object detection involves searching image components for regions matching the characteristics of target objects in a training database. This matching is achieved by comparing object templates to images across scales, rotations, and positions, with correlation techniques like Sum of Squared Differences (SSD) determining similarity. Recent advances have shown the responsiveness of image-based object detectors to training data.

B. Scope Of Research

The FDA (Food and Drug Administration) and OHRP (Office of Human Research Protections) are responsible for implementing security measures for human subjects. An image processing-based unmanned aerial vehicle (UAV) requires human involvement through the use of a joystick and camera. It will lessen the amount of time that workers spend on complicated tasks. Unmanned Aerial Vehicles (UAVs) occasionally employ expensive laser sensors and multisensor integrated systems to identify objects and individuals. This project will help to replace the laser sensor and utilize less expensive methods to service the site. This research seeks to compensate for non-detected object and processing face errors, as a UAV is a costly vehicle that cannot be lost.

C. Image Processing

Image processing refers to the procedure of converting an image into digital format and applying various adjustments to enhance the image or extract valuable information. This form of signal processing takes an image, such as a picture or video frame, as input and generates an image or related features as output. Image processing systems typically employ established signal processing techniques on images, treating them as twodimensional signals. The process generally encompasses three fundamental steps:

- Importing the image using digital photography or an optical scanner.
- Analyzing and modifying the image, including tasks like data compression, image enhancement, and identifying patterns in satellite photos that are imperceptible to the human eye.
- Outputting the final result, which involves generating a modified image or a report based on the analysis of the image.

1) Block Diagram

In Figure 4, the camera captures an image, and the algorithm determines features from this image. These features are then used to collect putative points. By utilizing these putative points, the object to be identified can be determined within the image.

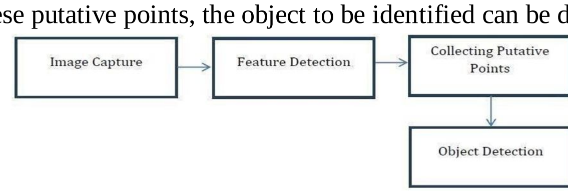


Figure 4: Diagram of object detection

2) Face Detection

Straightforward face tracking solution by breaking the tracking issue up into three distinct issues:

- Find a face to track
- Recognize features on a face to track
- Monitor countenance

a) Step 1: Face Detection for Tracking

The initial step involves the detection of a face to enable tracking, leveraging visual data. Utilizing a cascade object detector, one can pinpoint a face within a video frame. This detector utilizes the Viola-Jones detection technique, which employs mathematical models of Haar-like features and Viola & Jones algorithms (further elaboration on this will be provided later). Alongside a trained classification model, the cascade object detector can identify various objects, including faces, as per its default configuration. The resultant output signifies the detected face. Despite its capability to track a face across consecutive video frames, there is a risk of losing track if the face tilts or the subject turns their head. This limitation stems from the type of classification model utilized for detection. Due to the computational intensity of performing face identification for each video frame, this instance employs a basic facial feature for tracking to circumvent this challenge.

Figure 5: Detected Face

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

b) Step 2: Determine Which Face Features to Monitor

Finding a characteristic that will enable us to track the face comes next once we've found it in the video. We may utilize the color, texture, or form, for instance. We select an attribute that is exclusive to the thing and doesn't change even when it moves. In this instance, the attribute to track is color. The color does not alter as the face moves or rotates, giving a strong contrast among the face and the backdrop (fig. 3 is an output).

c) Step3: Track the Face

We may utilize the vision with the color set as an attribute to track. After identifying the face in the video as a place that was filled in the output area based on geometric coordinators, we are able to differentiate between the actual form of the face and the backdrop that corresponds to it.

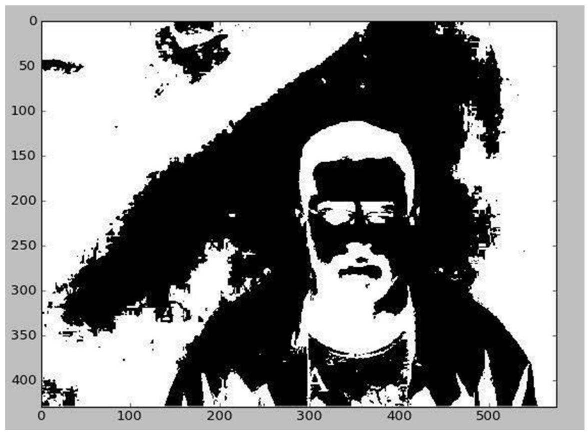
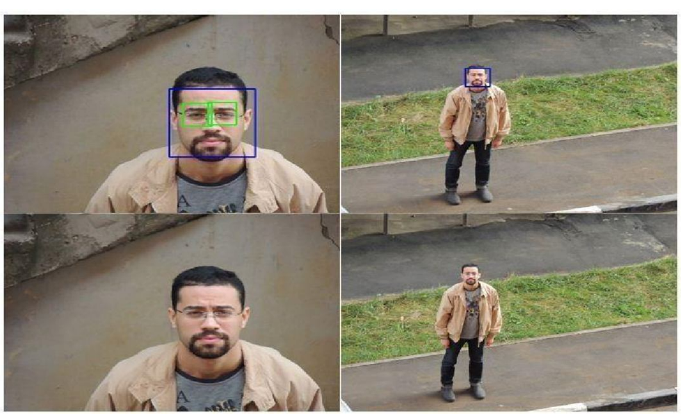


Figure 6: Face Tracking Features

D. Object Detection

We provide algorithms to display feature spaces utilized by object detectors in this and the preceding section. We discovered that these representations provide us fresh perspectives on object detection system failures and enable us to study the systems in novel ways. Thus, we are presenting an algorithm that uses form characteristics to recognize automobiles. An method for identifying a particular object based on locating the entire object (cars and human bodies) is shown in the following code. Even with little plan rotation or size changes, it is still able to identify things. Additionally, it can withstand a minimal degree of blockage and out-of-plane rotation. We attempted to identify objects in figures by analyzing their corresponding shapes. A parking notice and a nonfrontal complete human body are clearly represented in the Cascades are used in the detecting process. We require a tracking algorithm in order to track vehicles. The Haar Cascades' high rate of false illustration. Two primary things are also represented by the image on the left (see fig.): a tree and the complete frontal human body.

positives makes it an unsuitable option for vehicle tracking. What does the term "false positives" mean? Generally speaking, a false positive occurs when a condition that was tested for is inadvertently discovered to have been recognized during an evaluation procedure.



lets see The vehicles case Now. The easy wayto do vehicle Haar

***Figure 7:** Detected Objects*

526 rear-facing automobile photos (360x240 pixels, no scale) were used to train the Haar-Cascade cars3.xml (see fig.). The photos, which were shot of southern California's freeways, were obtained, which Brad Philip and Paul Updike [21–24] offered. Any moving item is identified as a car by this method. We may utilize the BGSLibrary [25] to preserve the foreground masks. The application of a BS algorithm is one suggestion. But, a blob tracker algo is required for vehicle tracking; a BS method alone is inadequate. The original YouTube video was taken by myself. To implement various vehicle monitoring and counting recommendations, we must:

- Do a background subtraction first.
- Forward the front mask to cvBlob.
- Several techniques for obtaining the track, ID, and centroid of moving objects are provided by the cvBlob library. Additionally, we may provide the locus & angle of the object tracked, as well as whether to build a bounding box.
- Examine our movie to see whether the center of the moving item has crossed any virtual lines or zones.

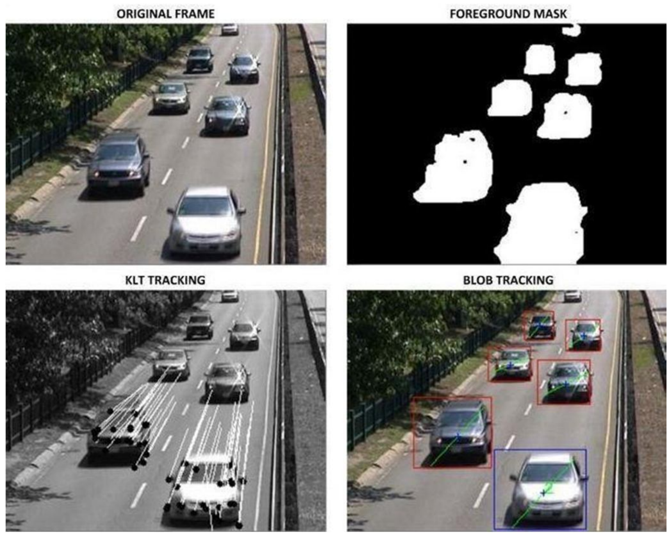


Figure 8: Detected Vehicles

E. Mathematical Modelling or Algorithms

1) Haar Like Features

The usage of Haar-like features in object detection [16-20] is proposed first time by Paul Viola and Michael Jones in Viola & Jones (2001). Subsequently, this approach garnered significant interest from other researchers, leading several to expand its application to additional object identification domains. The Haar Detection Cascade's principle is to remove negative samples with little processing. Several classifiers are calculated for each sub-region present in the picture. The picture is released and more processing is done if the subregion fails any of the classifiers. A subregion moves on to the next step, which calls for a little bit more processing, if it passes the first, which needs minimal work. The picture moves on to another subregion if it passes this classifier. For face detection to happen, a subregion of an image has to pass each of these classifiers. Although a classifier may be trained to increase in accuracy, OpenCV's face detection classifier serves our needs well.

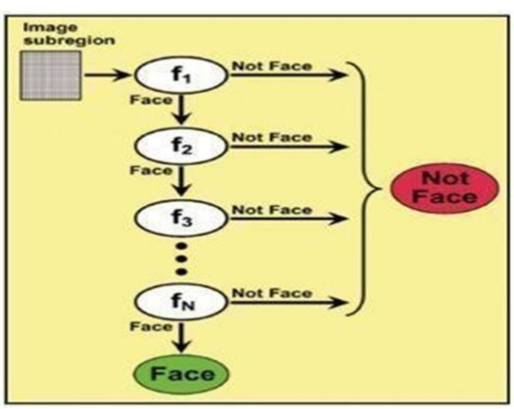


Figure 9: Haar Classifier

An extensive collection of two-dimensional (2D) Haar functions that may be used to represent an object's local appearance is known as a Haarlike feature [25]. These are made up of one or more rectangular sections surrounded by a template. The following equation yields the feature value f of a Haar-like feature with k rectangles:

$$f = \sum_{i=1}^{k} w^{(i)}.\mu^{(i)}$$

Where the ith rectangle around picture x represents the mean intensity of the pixels, and μ(i) is that value. The rectangle mean is denoted by the quantity μ. The weight allocated to the ith rectangle in (equation) is μ(i). Typically, default integer numbers are used to determine the weights of a Haar-like feature's rectangles so as to satisfy the following equation:

$$\sum_{i=1}^{k} w^{(i)} = 0$$

The prominence of Haar-like characteristics can be attributed, in part, to their extremely appealing trade between evaluation pace and preciseness. For the face identification task, Viola and Jones used a mere weak clssifier built on Haar-like features that cost only 60 CPU instructions, yielding a single wrong minuses & forty percent wrong pluses. The utilization of integral pictures, which, once calculated, may be quickly employed to assess any Haar-like feature at any scale in constant time, is primarily responsible for the high evaluation speed. Numerous distinct Haar-like characteristics have been reported in the literature [26–29] since Papageogiou et al. introduced horizontally and vertically aligned Haar-like features. The amount of rectangles and their orientation in relation to the template are the primary variations in the Haar-like features idea. In order to capture diagonal structures in an object's appearance, Jones and Viola proposed diagonal characteristics. Lienhart added effectively computable rotated Haar-like characteristics to the Viola and Jones features. They obtain a 10% reduction in false alarm rate for a given true positive rate using the enriched set. To define non-symmetrical characteristics of an object's appearance, Li et al. presented Haar like features with discontinuous rectangles.

2) *The Viola Jones Framework*

One of the earliest reliable real-time face detectors still in use today is the creation of Viola and Jones. The integral picture, the adaboost, and the attentional cascade structure are the three key concepts that underpin this detector's performance [30–31].

By combining progressively more complicated classifiers, the attentional cascade of classifiers enables the fast removal of background areas from the picture and increases computation on regions that appear to be objects. Boosters (low T) are utilized initially, which are simpler and thus faster, to reject most negative windows and pass nearly all positive ones. Next, to reject the significantly smaller number of challenging negative windows, boosted classifiers (with high T) that are more sophisticated and hence slower are employed.

F. Applications

Object detection is becoming more widely used in a variety of sectors, with applications ranging from office efficiency to personal security. Numerous fields in computer vision, such as, autonomous car systems, and inspection of machines, use object detection and recognition. There are still significant obstacles in the realm of object identification. The potential applications of object detection in the future are virtually limitless. We can talk about a few recent and upcoming applications here:

1) CHARACTER'S IDENTIFICATION
2) AUTOMATED CARS
3) OBJECT'S TRACKING
4) DETECTION OF FACES
5) EXTRACTION OF AN OBJECT FROM A PICTURE
6) DETECTING SMILING FACE
7) IDENTIFICATION OF AN ACTIVITY
8) DETECTION OF A PEDESTRIAN
9) WATERMAKRING DIGITALLY
10) PICTURING IN MEDICAL
11) TRACKING OF BALL IN A GAME

12) RECOGNITION OF AN OBJECT FROM IMAGE
13) PRODUCTION INDUSTRY
14) ATOMATED ROBOTS
15) CCTV CAMERAS
16) AUTOMATIC TARGET RECOGNITION
17) COUNTING OF OBJECTS
18) OBJECTS WHICH ARE AVAILABLE ONLINE

In computer vision, object detection and tracking find extensive uses in fields including robotics, medical imaging, augmented reality, and traffic monitoring. Medical image analysis is one of the many image processing applications that make extensive use of Bayesian classification techniques. To categorize the picture into a predetermined number of clasees, the basic process involves combining clinical information in previous terms with datadriven exposure in the probability terms. Tissue tracking and Brain MRI classifications are of primary importance [33]. Additionally, augmented reality provides a variety of tracking application domains; yet, the challenge of precise and real-time tracking remains unsolved, despite the field's daily growth. Amines is the most well-known application, operating on the virtual reality premise. Knowledge based augmented reality for maintenance assistance (KARMA)[34] and a mobile augmented reality systems for exploring the Urban Environment (MARS)[35] are two more similar systems that were created based on tracking augmented reality usage. It also supplies input for visual tasks at a higher level, such 3D representation and reconstruction. It is also crucial for content-based indexing and retrieval in video databases. This extremely moving object detecting technology is utilized to help disabled people navigate around their houses in today's technologically advanced environment.

IV. COMPARATIVE STUDY

The suggested approach is based on object detection, which helps with object prediction in a number of ways, including reducing accidents and helping people manage their daily schedules while keeping themselves safe from dangers or impediments. The accuracy of the suggested model is 72%. In the future, more datasets will be included to enable more accurate large-scale item prediction. [32] By Sunita Joshi, Neha Gupta, Mitali, and Gautam Yadav, "A Machine Learning Approached Model to Identify the Object for Visually Impaired Person (2023)" The goal of this project is to present an enhanced method for predicting items surrounding a visually impaired individual using a functional, real-time model.

Applications of object recognition that have been thoroughly researched include computer systems in cars, facial and character recognition, and more. There have been multiple uses for object detection, such as monitoring and searching. In order to increase the effectiveness and precision of object identification, this research presents certain fundamental ideas that are utilized in conjunction with the Python 2.7 OpenCV package. [36] Object recognition is a well-known computer technique linked to computer vision and image processing that is used to identify particular classes (people, flowers, animals, etc.) in digital photos and videos, as suggested by Bumika, Gupta (2017) et al. Its main goal is to identify things or instances of them. A system requires many components to carry out the object detection task in order to identify objects in photos or videos. These are the verifier, hypothesis, feature detector, and model database. An overview of the many methods for object identification, object location, object classification, feature extraction, information extraction from images and videos, and other tasks is given in this white paper. Remarks are derived from the studied literature, and significant problems with object detection are also pointed out. New researchers in the field of object identification are assisted by the provision of source code and online dataset information. There are also suggestions for potential multiclass object identification systems given. Researchers who are fresh to the field can benefit from this study. [37] Kartik Umesh Sharma published "A review and an approach for object detection in images" in 2017. An object identification system that can identify actual items in digital photos and movies was proposed by Nileshsingh Thakur. Any class of object, such as B. People, automobiles, etc., can include objects.

“A Review of Detection and Tracking of Object from Image and Video Sequences” (2017) by Mukesh Tiwari and Dr. Rakesh Singhai focuses on Because objects frequently change in motion and vary in size, object recognition and tracking is one of the most important fields of research. One of the main areas of research includes scene size changes, occlusions, appearance changes, ego movements, and lighting changes. shown as a single entity. 38 Feature selection is particularly crucial to object tracking. Numerous real-time applications, like video surveillance and vehicle recognition, can benefit from this. Tracking makes reference to an object's appearance and movement in order to address the detection challenge. The majority of the algorithms are devoted to tracking techniques for video sequence smoothing. However, very few techniques make advantage of previously discovered details on the form, color, texture, etc. of an item.

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 II Feb 2024- Available at www.ijraset.com

In this paper, we incorporate the aforementioned object parameters into a tracking algorithm and explain and evaluate it. This white paper's goal is to examine and evaluate earlier methods for object tracking and detection utilizing different phases of video sequences. Additionally, we point out holes and suggest the new strategy for betterment of object tracking in frames.

"A Novel Machine Learning Method for Identifying and Identifying Objects," as suggested by Aishwarya Sarkale (2018). This implies that visual item identification is a well developed human capacity. However, object recognition becomes problematic for machines. Neural networks were first presented to computer science in this way. "Artificial neural networks" is another term for neural networks. Computer representations of the brain called artificial neural networks aid in item recognition and identification. This white paper explains and illustrates the accuracy of several neural network types, including ANN, KNN, FASTER R-CNN, 3DCNN, and RNN. We may infer that ANN offers the greatest accuracy for object detection in certain test situations based on investigations in numerous research publications that compare and discuss the accuracies of various neural networks.

While other algorithms, like Convolutional Neural Networks and Fast Convolutional Neural Networks, only partially explore the image, YOLO uses convolutional networks to predict bounding boxes and the class probabilities of those boxes to fully examine the image. Identify photos more quickly than with previous methods. [40] Gitapriya S. et al. (2019). Detecting things with a You Only Look Once (YOLO) strategy is the suggested objective. Comparing this approach to other object identification methods, there are various advantages.

According to R Sujeetha (2020), object identification and tracking has the potential to become into one of computer vision's most significant and active subfields, if not the most widely used. [41] Researchers have created several simplified and specialized techniques as a result of its widespread usage in security tracking modules, government monitoring, and many other applications. Nevertheless, there is an issue with realtime object tracking and detection methods. This work is complicated by the need for real-time tracking, well-optimized results, dynamic computing to identify efficient performance of temporal components, and tracking of various objects. Although a number of strategies have been established, more can be done. In this technique, we label the identified layers and make sure their correctness can be recognized and simulated in real-time by extra external hardware by utilizing the CNN algorithm, TensorFlow and OpenCV libraries. Lastly, we see effective and finely tuned object tracking and detection systems.

"Deep Learning-Based Object Recognition for Visually Impaired Individuals" (2018) by Rebeiro Sharlene Huda Noordean and Sara Carlton1 concentrate on the development of object detection methods. Deep learning classifiers use convolutional networks instead of handengineered features, starting with the former. Additionally, the dataset that is frequently used to train, test, and validate the item that the new models identify is presented. Additionally provided is a broad review of the suggested system and how it will be used to provide object identification for the blind. There is also a brief description of the general difficulties encountered in object detection. [42]

“Object Detection System for Visually Impaired(2021)” by Dr K Sreenivasulua , P. Kiran Raob , Dr. VenkataRamana Motupallic [43] emphasizes the design of several For the purposes of this project, a single class vector support machine classification method proved to be quite effective since it provided significant adaptability by assigning one SVM to each category. This manner, without relying on the rest of the system, each of them—new and experienced—can complete their training alone. Along with allowing each class to develop at its own rate, this architecture also ensures that there is no imbalance in the quantity of examples for each category, which could compromise the classifier's quality if a different type of architecture, like One Vs. All, were chosen. For each of the system's classes to be trained, an automated grid search is essential. Because the SVMs' ideal parameter values vary depending on the class. This, together with the construction of numerous SVMs—one for each class—allows the system to be fully autonomous and capable of ongoing improvement through the use of user-provided data. This continual development includes both the capacity to learn new classes that users supply and the increasing accuracy in already-existing classes as your training data expands. The first goal was to create a categorization approach that took the items' context and similarity into account. Throughout the project's development, Including AlexNet's pre-trained network allowed for the conclusion that all of the items that comprised the database could be made in a generic way. This can be attributed to the previously described network's training using one million photos split over a thousand distinct categories, which prevented objects from being grouped based on context and similarity. That's why it was also decided to replace the way a group of programs that facilitate user-server communication are implemented with a single application that can identify any kind of item.

“Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People.(2016)” by Samruddhi Deshpande and Ms. Revati Shriram , An innovative text detection and recognition technique is presented in this paper. The technique explains how to assist blind persons by having printed letters read from things. This suggested approach solves the reading difficulty for blind people. This system identify words in picture and separate the item from the recorded image. The suggested technique for text localization in real-world photos is shown.

The text content of the image is detected using MSER. By doing this, text patterns from various backgrounds will be extracted. This is the text detection method that is resilient. On the localized text patterns, OCR is employed for text recognition. For blind users, the system will convert the translated text to an audio output. The findings show that great performance in text detection and identification may be attained with MSER and OCR. [44]

V. FUTURE SCOPE

1) Enhanced Accuracy and Reliability Continued research and development to improve the accuracy and reliability of object detection algorithms. This includes reducing false positives and negatives, especially in complex or dynamic environments.
2) Real-time Processing Focus on achieving real-time processing capabilities to provide instantaneous feedback to users, allowing them to navigate their surroundings more efficiently and safely.
3) Semantic Understanding Progress in incorporating semantic understanding into object detection algorithms. This involves not only recognizing objects but also understanding their context and potential impact on the user's safety and navigation.
4) Multi-Modal Feedback Implementing multi-modal feedback systems that combine audio, tactile, and visual cues to provide a more comprehensive understanding of the environment. This can enhance the user's spatial awareness.
5) Edge Computing Exploration of edge computing solutions to process data locally on the device, reducing dependence on cloud services and ensuring faster response times, especially in areas with limited network connectivity.
6) Adaptability to Various Environment Customization of object detection models to adapt to different environments and scenarios, including indoor and outdoor settings, crowded spaces, and diverse lighting conditions.
7) User-Friendly Interface Development of user-friendly interfaces that allow individuals to easily interact with the system, customize preferences, and receive information in a way that suits their needs and preferences.
8) Machine Learning Advancements Leveraging advancements in machine learning, such as active learning and continual learning, to improve the adaptability of the system over time and enhance its ability to recognize new objects or adapt to changing environments.
9) Collaboration with Accessibility Advocates Collaboration with blind and visually impaired communities to ensure that the technology aligns with their needs, preferences, and feedback. Involving end-users in the design and testing phases can lead to more effective and user-friendly solutions.
10) Global Accessibility Standards Advocacy for and adherence to global accessibility standards to ensure that object detection technologies for the blind are widely adopted and meet the needs of users worldwide

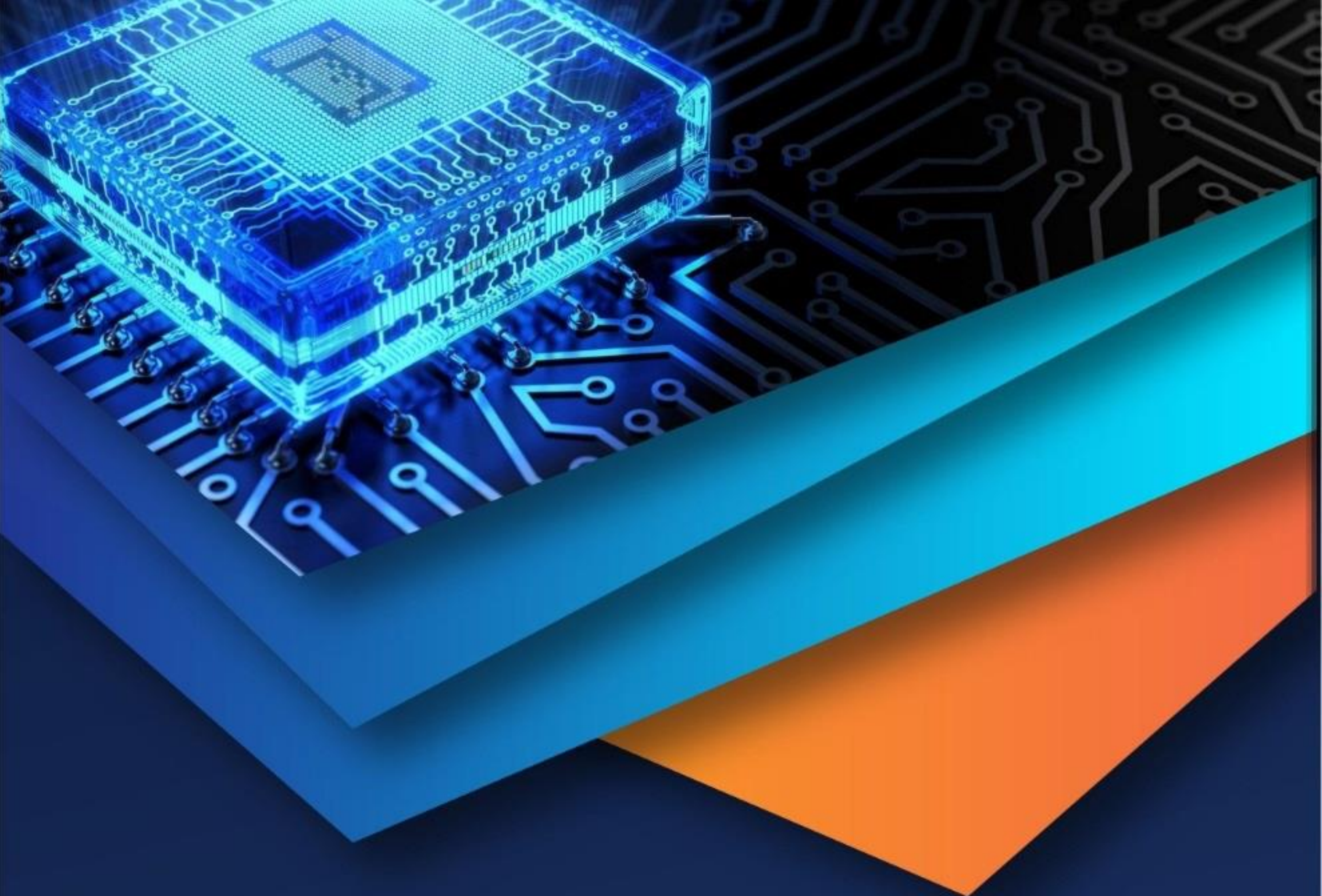
VI. CONCLUSION

This post was created from the ground up, covering everything from the necessity of computer vision to the methods and rationale behind object and face detection. We went into great depth on every idea related to object and face detection, as well as the significance of this subject. Our findings demonstrated that the primary goal was object detection, and the output items were identified from the actual picture. In surveillance and other contexts, the face detection software can be used to identify and track individuals. To help understand the effectiveness of image detecting structures, we presented a tool. We introduce strategies for visualizing object detector success spaces. The findings of Paul Viola and Michael Jones (2001) corroborate ours. In order to validate findings, we also presented and assessed a number of techniques for visualizing object identification attributes.

REFERENCES

[1] Wu, R.B. Study on Face Recognition and Intelligent Video Surveillance Applications in Prison Security. 2019; 6: 16–19; China Security Technology and Application.
[2] Research and Challenge of Deep Learning Methods for Medical Image Analysis by Tian, J.X., Liu, G.C., Gu, S.S., Ju, Z.J., Liu, J.G., Gu, D.D. In 2018, Acta Automatica Sinica 44: 401-424.
[3] The research status and development trend of intelligent detection technology and industrial robot target identification are presented by Jiang, S.Z., and Bai, X. 2020; 36: 65– 66 Guangxi Journal of Light Industry.
[4] Deep Convolutional Neural Networks for ImageNet Classification by A. Krizhevsky, I. Sutskever, and G. Hinton. 2012; 25: 1097–1105, Advances in Neural Information Processing Systems.
[5] ImageNet Large Scale Visual Recognition Challenge / Russakovsky, O., Deng, J., Su, H., et al. 2015; 115: 211-252 in International Journal of Computer Vision.
[6] Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation Girshick, R., Donahue, J., Darrel, T., Malik, J. In: Columbus, 2014, Computer Vision and Pattern Recognition, pp. 580–587.
[7] Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition by He, K.M., Zhang, X.Y., Ren, S.Q., and Sun, J. 2015; 37: 1904–1916; IEEE Transactions on Pattern Analysis & Machine Intelligence.

[8] Fast R-CNN by R. Girshick is published in the IEEE International Conference on Computer Vision Proceedings. 2015 Santiago, pp. 1440–1448.
[9] Faster R-CNN: towards real-time object identification using region proposal networks, Ren, S.Q., He, K.M., Girshick, R., Sun, J. In: Neural Information Processing Systems Advances. Montre al, 2016; pages 91–99.
[10] Farhadi, A., Grishick, R., Divvala, S., and Redmon, J. Unified, Real-Time Object Detection—You Only Need to Look Once. In: Las Vegas, 2016; Computer Vision and Pattern Recognition, pp. 779–788.
[11] Redmon, J. and Farhadi, A. Yolo9000: enhanced, accelerated, and fortified. In: Hawaii.2017, Computer Vision and Pattern Recognition, pp. 7263-7271.
[12] Farhadi, A. and Redmon, J. (2018) Yolov3: A step at a time. ArXiv: Pattern Recognition and Computer Vision.
[13] SSD: Single Shot MultiBox Detector Liu, W., Anguelov, D., Erhan, D., et al. 2016, pp. 2137, European Conference on Computer Vision.
[14] YOLOv4: Ideal Speed and Accuracy of Object Detection, Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. 2020 ArXiv: Pattern Recognition and Computer Vision
[15] [Online] Available: scholar.google.fr/scholar?hl = fr&q = Object + detection + using + Haar cascade + Classifier&btnG = &lr =
[16] [Online] Available: lab.cntl.kyutech.ac.jp/ kobalab/ nishida/opencv/OpenCVObjectDetection HowT o.pdf
[17] [Online] Available: cs.colby.edu/maxwell/courses/cs397 vision/F 07/papers/viola − Faces − cvpr01.pdf
[18] [Online] Available: cbcl.mit.edu/publications/ps/heisele x3hei.lo.pdf
[19] [Online] Available: arxiv.org/pdf/1502.05461v1.pdf
[20] [Online]Available:www.researchgate.net/profile/VitorSantos6/publication/267868 2 Automatic Detection of Cars in Real Roads using Haar−like Features/links/552c0bae0cf 2e089a3ac3bc3.pdf
[21] [Online] Available: www.svcl.ucsd.edu/projects/traffic/
[22] [Online] Available: github.com/andrewssobral/bgslibrary#bgslibrary
[23] [Online] Available: github.com/andrewssobral/simple vehicle counting
[24] [Online] Available: github.com/andrewssobral/bgslibrary
[25] C.P. Papageorgiou, M. Oren, T. Poggio, A general framework for object detection, in: ICCV ’98: Proceedings of the Interna- tional Conference on Computer Vision, Washington, DC, USA, 1998, pp. 555-562.
[26] S.Z. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, H. Shum, Statistical learning of multiview face detection, in: ECCV ’02: Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Sciences, vol. 2353, London, UK, 2002, pp. 67-81.
[27] R. Lienhart, J. Maydt, An extended set of Haar-like features for rapid object detection, in:
[28] ICIP ’02: Proceedings of the International Conference on Image Processing, 2002, pp. 900- 903
[29] P. Viola, M.J. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR ’01: Proceedings of the Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, 2001, pp. 511-518.
[30] M. Jones, P. Viola, Fast multi-view face detection, Technical Report MERLTR2003-96, Mitsubishi Electric Research Labora- tories, July 2003.
[31] Viola, P.; Jones, M., ”Rapid object detection using a boosted cascade of simple features,” Computer Vision and Pattern Recog- nition (CVPR), 2001 IEEE Conference on, June 2001.
[32] [Online]Available:ir.library.louisville.edu/cgi/viewcontent.cgi?a rticle = 2731&context = etd
[33] “A Machine Learning Approached Model to Identify the Object for Visually Impaired Person(2023)”by Sunita Joshi ,Neha Gupta, Mitali, Gautam Yadav
[34] John Melonakosa and Yi Gaoa and Allen Tannenbauma, Tissue Tracking: Applications for Brain MRI Classification , Georgia Institute of Technology,14 Ferst Dr, Atlanta, GA, USA
[35] KARMA, Computer
[35] Feiner, S., MacIntyre, B., Höllerer, T., and Webster, T., A touring machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment, In Proc. ISWC '97
(Int. Symposium on Wearable Computers), Cambridge MA, (Oct 1997)
[36] Object recognition, as proposed by Bumika, Gupta (2017)
[37] “A review and an approach for object detection in images” by Kartik Umesh Sharma Nileshsingh Thakur in 2017.
[38] “A Review of Detection and Tracking of Object from Image and Video Sequences” (2017) by Mukesh Tiwari and Dr. Rakesh Singhai
[39] “An Innovative Machine Learning Approach for Object Detection and Recognition”As Proposed by Aishwarya Sarkale (2018).
[40] Gitapriya. S (2019) et al. The proposed goal is to detect objects using a You Only Look Once (YOLO) approach.
[41] R Sujeetha (2020) the proposed object detection and tracking could become a huge and active field of computer vision, if not definitive and prevalent.
[42] “An Object Recognition for Blind People Based on Deep Learning (2018)” by Rebeiro Sharlene Sara Carlton1, Huda Noordean
[43] “Object Detection System for Visually Impaired(2021)” by Dr K Sreenivasulua , P. Kiran Raob , Dr. VenkataRamana Motupallic
[44] Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People.(2016)” by Samruddhi Deshpande and Ms. Revati Shriram



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY