iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55817

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 11 Issue IX Sep 2023- Available at www.ijraset.com

Multi Object Detection

Geetha Siva Srinivas Gollapalli[1], Yaswanth Chowdary Thotakura[2], Shalom Raja Kasim[3], Kalyan Kumar Doppalapudi[4]
[1, 2, 3, 4]*Department of CSE, Bachelor of Scholars, SRM University-AP*

Abstract: ***In the realm of computer vision, the ability to accurately detect and comprehend objects within images and videos is of paramount importance. This research is dedicated to advancing the field of object detection, a critical component of computer vision, with a particular focus on leveraging Convolutional Neural Networks (CNNs) to enhance accuracy. CNNs have revolutionized object recognition tasks, outperforming traditional methods like Viola-Jones, SIFT, and HOG. The study explores the underlying architecture of CNNs, elucidating how convolution, pooling, and flattening layers enable efficient image processing and object identification. Object detection holds immense practical significance, spanning applications such as autonomous vehicles, surveillance, and medical imaging. By delving into the intricacies of CNNs and their role in object detection, this research contributes to the ongoing evolution of computer vision, promising advancements in diverse sectors of industry and technology.***

Keywords: Object detection, Object tracking, Convolutional neural network (CNN), Video Analysis, Deep Learning

I. INTRODUCTION

The branch of study known as "computer vision" (CV) aims to create methods that will enable computers to "see" and comprehend the content of digital images like pictures and movies. Computer vision systems must be able to identify the current objects and all of their attributes, including sizes, forms, colors, textures, and spatial arrangements. The objective is to explain and characterize pictures. For instance, noise reduction, object detection, etc. Compared to its allied subjects, such as, computer vision does significantly more. processing images, machine vision. In the field of image processing, we treat photos as squares of colors or shades, much like how you would read a double-dimensional matrix. This method enables us to apply arithmetic and procedures to sharpen images, locate edges, or identify things within them. It's similar to utilizing a map to interpret and enhance what is in a picture In artificial intelligence, webcams serve as the equivalent of pupils, taking photographs or movies of what they "see" in the environment. Following that, to assess particular characteristics such hues, shapes, or product flaws, these photographs are processed using special technology, including lenses and computer algorithms. Such information is beneficial in many applications, such as determining whether goods are correctly manufactured in a plant.

Simply put, image recognition is a computer technique that enables computers to comprehend and identify objects in images. It's similar to teaching a computer to recognize and locate specific items in pictures or images, such as dogs, cats, or cars. Self-driving cars employ this technology to help them "see" the road and other cars, as well as in smartphone apps that can tell you what's in a photo. Object detection classifies numerous objects in an image and uses bounding boxes to show where each one is located. In other words, it's an adaptation of picture classification that includes localization tasks for a variety of objects. The objective of object identification, which differs slightly from object detection in this context, is to locate occurrences of a certain object in photographs. It is not about categorizing an image; rather, it is about figuring out whether or not an object exists in an image and if so, where exactly. By enclosing an object in a bounding box, object localization techniques locate the object in an image and determine its location. Differentiating many objects (instances) from the same class (each person in a group) is what instance segmentation does. It might be considered the stage after object detection. It involves not only identifying objects in a picture but also accurately producing a mask for each one that is found. A crucial area of computer vision is object tracking. It entails tracking an input object—which could be a person, a ball, or a car—across a number of frames.

The goal of artificial seeing, who attempts to make it possible for computers to comprehend and interpret digital images and videos, is introduced at the beginning of the written article. It highlights how crucial identifying objects is to robot vision.

II. PROPOSED WORK

A. Multi Object Detection

Multi-object monitoring aims at collecting the shifting position of several entities throughout the footage stream. In practice, multi-object tracking is always preceded by object detection and hence, tracking precision depends on object detection accuracy. Multi-object tracking finds use in Autonomous vehicles, Security and surveillance, Traffic control.

B. Different Heterogeneous Object Acquisition Styles and Methods

Researchers utilized several algorithms on computers for finding items in photographs prior neural networks gained popularity around 2013. They employed arithmetic to determine whether certain patterns indicated the presence of a substance by searching for particular motifs or properties, such as lines or hues. The names of these methods were Viola-Jones, SIFT, and HOG. In order to categorize sets of related traits and determine if they matched an object, they also used computer programs. Although a few of these approaches were effective, they were not without certain drawbacks. As a result, supervised learning, a sort of algorithm for learning, gained popularity since it was able to acquire information out of data and had the potential to be more precise. Now-used deep learning-based techniques outperform them by a wide margin. Utilizing neural network architectures such as RetinaNet, YOLO (You Only Look Once), CenterNet, SSD (Single Shot Multi-box detector), and Region proposals (R-CNN, Fast-RCNN, Faster RCNN, Cascade R-CNN), deep learning-based algorithms determine the labels of the objects based on their attributes.

C. Object Detection Applications

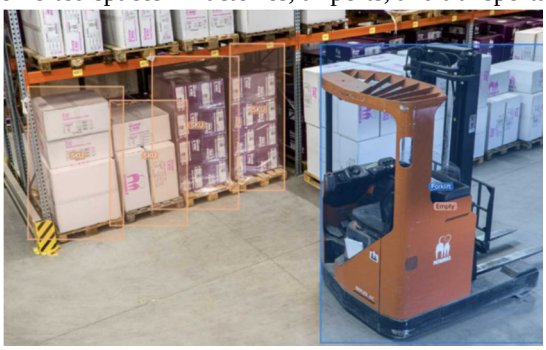
Within robotics, finding and identifying objects, such as people or cars, in images or videos is done by instructing the machine in question to do so. It enables machines to comprehend whatever is contained in an area of an image or footage, making it helpful in a variety of fields including spying and autonomous automobiles. Due to its capacity to automate and improve processes requiring object recognition, this technology has a wide range of applications across numerous sectors.

D. Face and Person Detection

Most facial analysis algorithms are powered by the identification of things. Something is frequently utilized for identifying an individual. within a group by detecting faces, classifying attitudes as well as symptoms, and submitting the generated container to a machine for collecting images. When you use your face to unlock your phone, you are probably already using one among the highest frequent application scenarios for identifiers: facial recognition. Discovery for individuals is frequently used to measure social distance or to count the number of people at retail establishments.

E. Intelligent Video Analytics

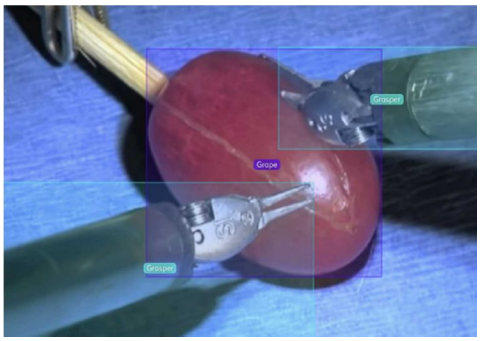
In order to understand how customers interact with products in retail settings where CCTV cameras are present, intelligent video analytics (IVA) uses object detection. These video streams are anonymized using a pipeline that blurs faces and removes identifiers from them. Some IVA use cases protect privacy by limiting camera placement to below-knee height and ensuring the system merely records a person's presence without having to physically examine their distinguishing characteristics. IVA is frequently utilized for recording wait times and gain entry to prohibited spaces in factories, airports, and transportation hubs.



F. Autonomous Vehicles

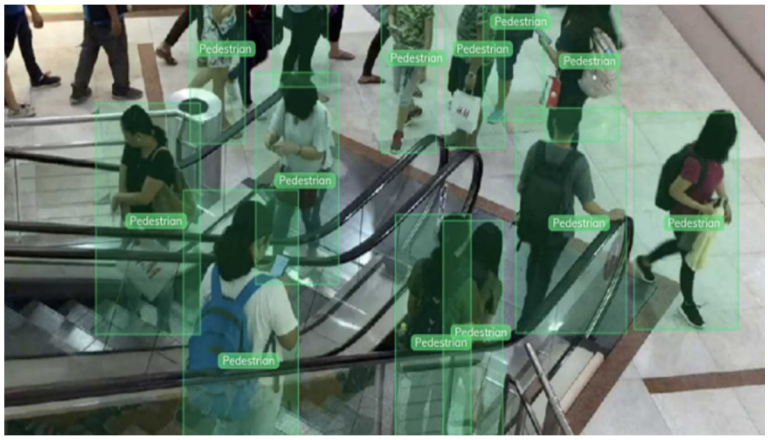
In order to navigate the road safely, self-driving cars use object detection to identify pedestrians, other vehicles, and barriers. LIDAR-equipped autonomous vehicles will occasionally use 3D object detection, which applies cuboids around things.

Tapes captured from endoscopes during critical operations is relatively noisy data. Using object detection, a surgeon can quickly identify difficult-to-see objects like polyps or lesions. Additionally, it is utilised to update hospital workers on the progress of the procedure.



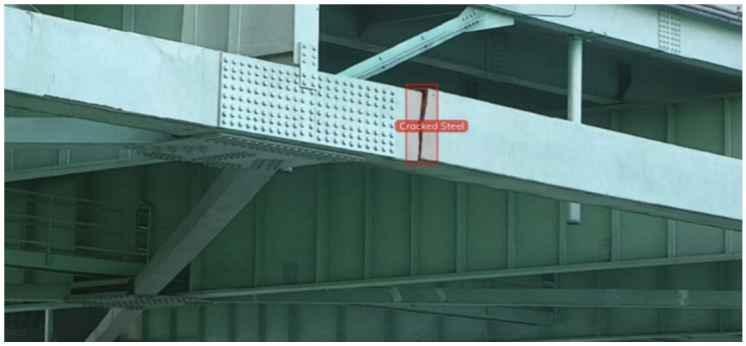
G. Pedestrian Detection

Robotics, video surveillance, and vehicle safety all use it, making it among the foremost important problems in visual analysis. The fundamental data needed to the semantic interpretation of video footage is provided by pedestrian detection, which is a crucial component of object detection research. However — This technology still confronts difficulties despite its generally great performance, such as the existence of occluding items or a variety of garment styles that affect the present-day sensors' precision..



H. Robot Guidance Using Intelligence

Unmanned aerial vehicles are capable of evaluating terrain with incredible sights anything they come across by using models stored in the cloud. For instance, they can be used to evaluate power lines or difficult-to-reach parts of bridges for cracks and other structural issues, replacing risky frequent helicopter operations.



III. METHODOLOGY

A. Convolutional Neural Network (CNN)

1) Convolution

Another name for a feature detector is a kernel or filter. Convolved feature and activation map are other names for the term "feature map." Finding the features in an image is the goal of a feature detector. If the pattern on the feature detector matches the corresponding portion of the input image, we obtain the maximum value for the feature map.

2) Pooling

Convolutional neural networks, also known as CNNs, filtering procedures are essential in reducing or streamlining the data that is retrieved from pictures. The maximum pooling technique prioritizes every important feature present while working with a grid of numbers denoting image attributes. It chooses the highest number in a limited location. A finer depiction of the characteristics is produced by average pooling, which determines the mean frequency for the area. Total Pooling, a fewer popular yet nevertheless valuable technique, adds together both the data throughout the region and provides perception in the general importance of attributes. Such pooling methods act as restrictions, bringing down the amount of detail of the material and making it easier for the machine learning algorithms to handle.

3) Flattening

In the world of neural networks using convolution (CNNs), flattening 2-dimensional arrays of feature maps into a single, continuous linear vector is crucial. Following the gathering of key picture elements using pools and convolutional layers, this process takes place. We establish a representation that is compatible with later artificial neural network layers, particularly layers that are fully linked, by transforming the attributes of visualizations into an exponential structure. The order of significance and linkages amongst the characteristics are preserved inside the matrix, but the geographic organization among the information disappears throughout the change. Through this process, info's complexity is substantially reduced, rendering it easier to operate in terms of processing and memory.

The flattened vector, which serves as a critical building block for activities like classification of pictures, object identification, and many artificial intelligence programs, contains the basic characteristics gathered from the source imagery. Essentially simple terms, smoothing fills the discrepancy separating the artificial network's capacity for successful retinal patterns identification and interpretation and the structure of the with feature extraction procedure.

4) ANN (Full Connection)

Completely Coupled (also known as FC) tiers function as an essential connection among the tiered feature retrieval stages with the ultimate choice-making procedure in the structure of a Convolutional Neural Network (CNN). Through attaching each cell with the ones in the levels above and below, these stages, which also incorporate weights and prejudices, provide worldwide interaction as opposed to the pools and convolution layers, which only have interpersonal relationships between neurons. FC layers are in charge of collecting and flattening the top-level characteristics retrieved from the preceding layers into a 1-dimensional matrix. They are often located near the end of the CNN. The traits of this matrix are then transformed into a format appropriate for the particular task, which can be image classification, object detection, or different machine vision utilization, by means of standard deep neural network layers, enabling complicated recognizing trends, culminates in the result of the stage of the algorithm for forecasts or choices.

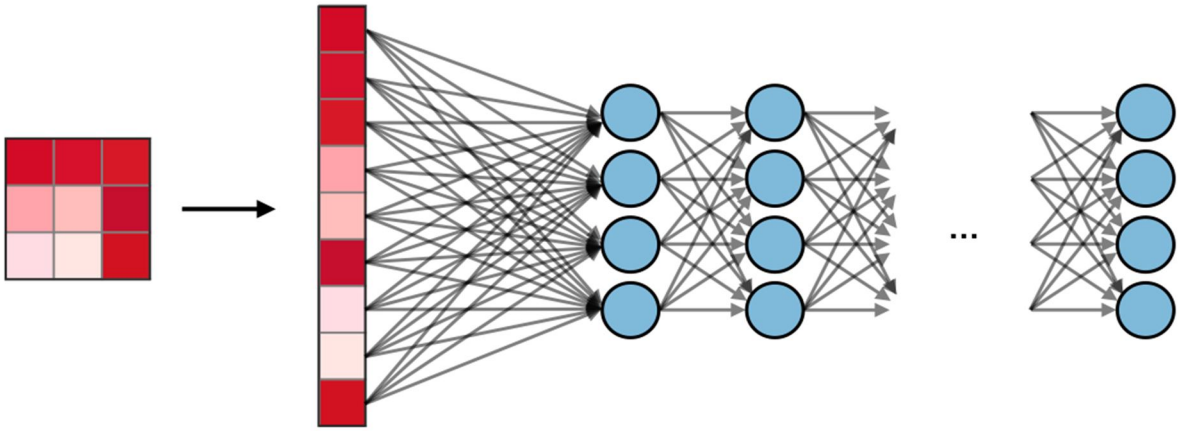


Fig. 1 Fully Connected (FC)

B. CNN Architecture

The three primary layers of a convolutional neural network are:

1) The Convolutional Layers
2) The Pooling Layers
3) The Fully Connected Layers

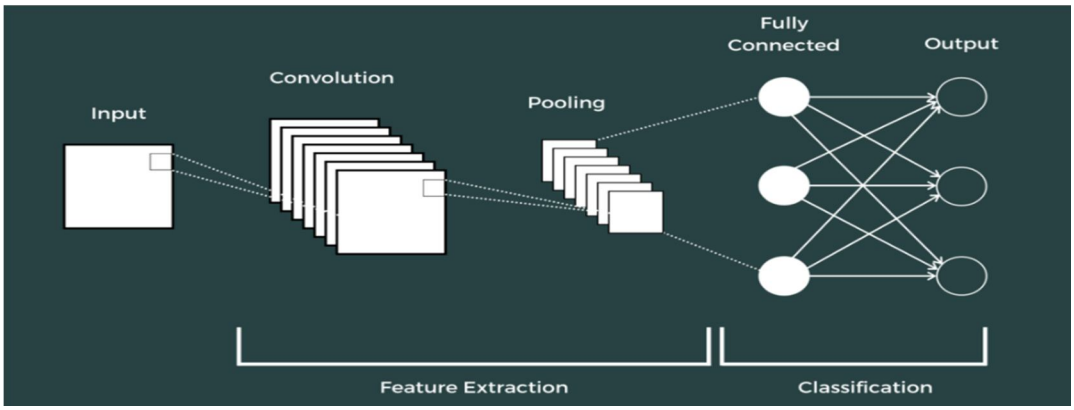


Fig. 2 CNN architecture

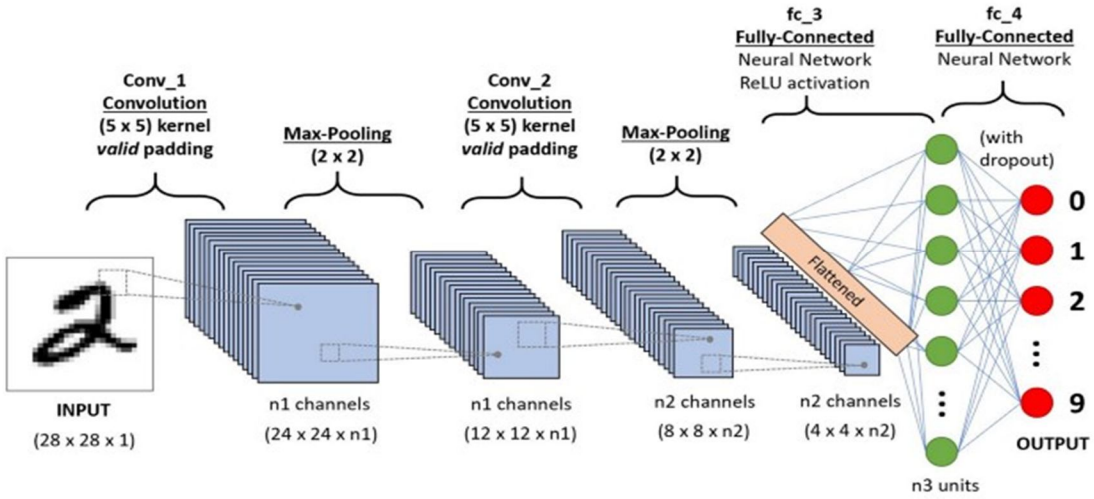


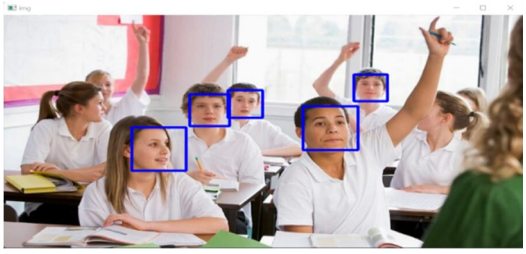
Fig. 3 CNN sequence to classify handwritten digits

IV. EXPERIMENTAL RESULTS

A. Face Detection

Typically, the visuals we see are in the RGB (Red, Green, and Blue) channel format. An RGB image is commonly stored in the BGR (Blue, Green, and Red) channel when it is read by OpenCV. For picture identification, we must transform this BGR channel into a grey channel. Grey channel is simple to compute and computationally less demanding because it only has one black-and-white channel. We will supply the following inputs to this cascade function:

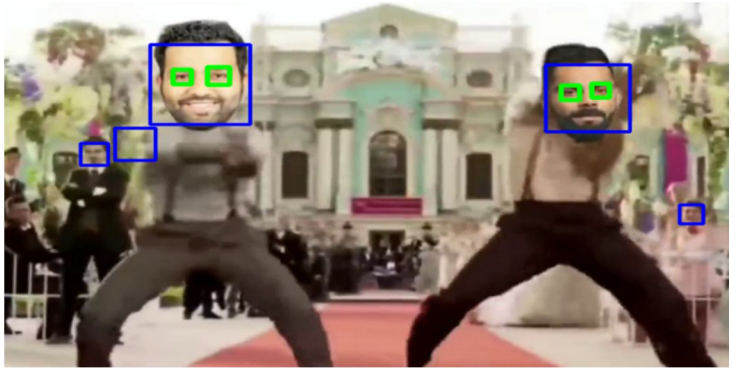
1) *A parameter for grayscale measurement*: within our predicament, drab
2) *Scale Factor:* At each level of scaling, the scale factor defines how much the image size is shrunk or expanded. The model's XML setup specifies a specified size for recognition during training, thereby building a scale pyramid. This suggests that at that particular size, the program can recognize faces. A larger face can, however, appear smaller by resizing the input image, which will help the algorithm recognize it. The likelihood of matching the model for detection is increased by using a reasonable estimate of 1.05, which denotes a modest 5% resizing increment. It's crucial to remember that this strategy may cause the algorithm to run more slowly due to higher thoroughness, leading to longer processing times. A scale factor like 1.4 can be used for quicker detection, the possibility of completely missing certain faces. A scale factor of 1.1 was used for our example because it found the proper balance in the particular context of the image being used.
3) *Min Neighbours:* a setting indicating the bare minimum every potential box needs companions to be present to be selected and kept. This setting will have an effect on how well the faces are discovered. Not many, greater-standard identifications are made when the worth is greater. At 3–4, it's an outstanding bargain. In our case, I determined that four neighbors would be the absolute minimum, which was perfect for the image I selected.
4) *Result:*

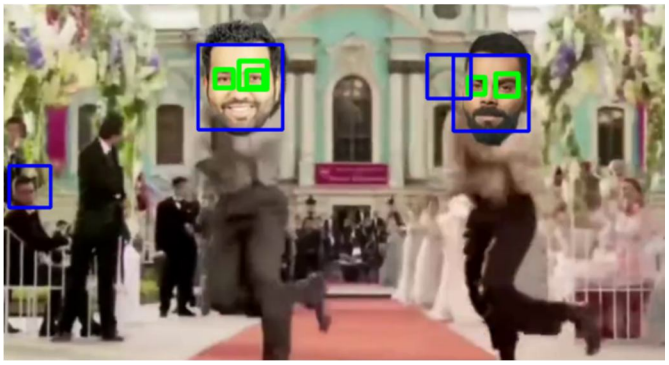


B. Face and Eye Detection

We have also included the haarcascade_eye.xml file to enable eye recognition. Using the videocap option, we've integrated video input. After obtaining the x-coordinate, y-coordinate, width (w), and height (h) of the detected face features through the detectMultiScale function, we proceed to create two numpy arrays, namely roi gray and roi color. The variable "gray" serves as the basis for building the numpy array roi gray, specifically used to extract the eye features (x, y, h, and w) and pass them to the detectMultiScale method. Subsequently, we iterate through the extracted face features (x, y, w, h), employing the numpy array roi color to generate rectangles. It's essential to note that roi color represents the array for the original RGB-scale image, while roi gray corresponds to the grayscale version utilized for efficient processing during dimension and coordinate extraction. Consequently, roi color is the appropriate choice when passing these coordinates.

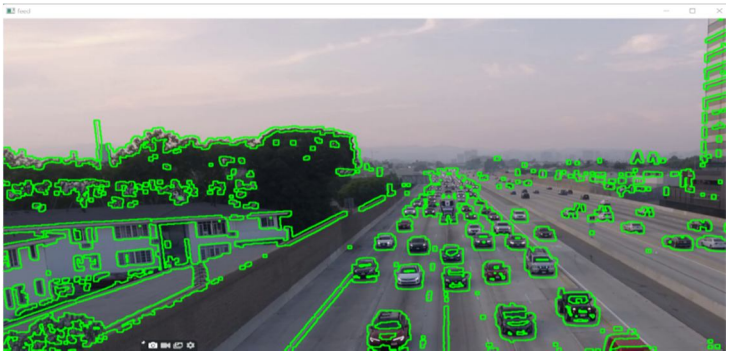
Result:

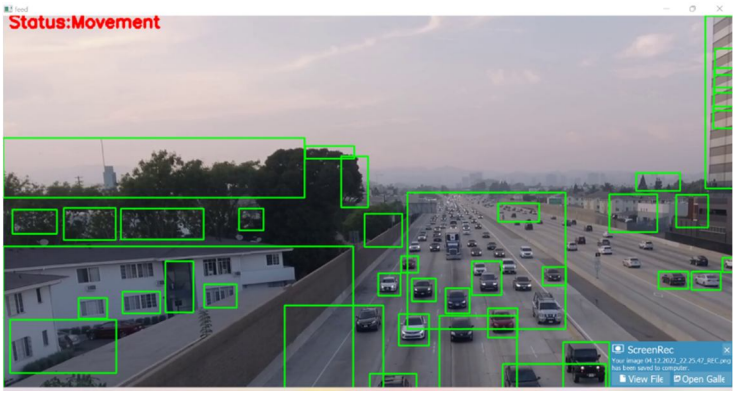




C. Vehicle Detection from Streaming Video

Having obtained the tuple (ret, frame), the next step involves converting the image from the BGR color channel to grayscale. This conversion to grayscale is undertaken to enhance overall processing efficiency, and subsequently, the classifier function "detectMultiScale" is deployed to extract crucial attributes: the x-coordinate, y-coordinate, width (w), and height (h). Moving forward, we iterate through each frame of the image, and for each frame, we generate a corresponding rectangle. This rectangle is established based on the previously extracted attributes and dimensions related to the cars within the image.



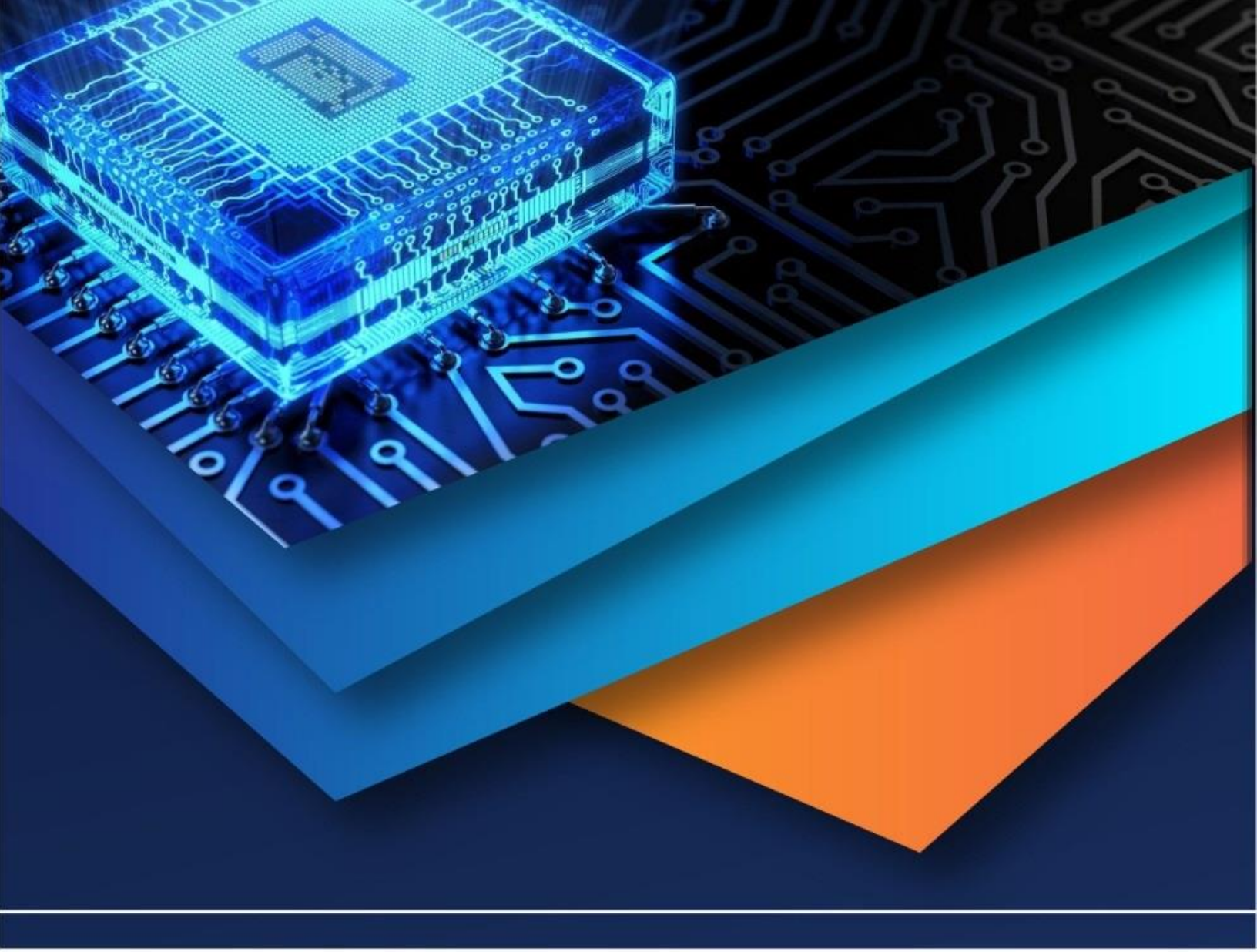


V. CONCLUSION

An essential component of computer vision, improving computer systems' recognition and comprehension of objects in photos and videos is the main goal of this study. The project attempts to increase the precision of object detection using convolutional neural networks (CNNs), which is crucial for applications like spying and autonomous vehicles. Simpler approaches like Viola-Jones, the SIFT algorithm, and HOG had been implemented prior to the development of these cutting-edge algorithms, but CNNs, or CNN have shown to be substantially more efficient. Numerous practical uses for identifying objects exist today, such as recognizing faces and monitoring of objects. The study explores the intricate scientific characteristics of CNNs, demonstrating how numerous axons and calculations enable visual analysis and recognition of items, potentially resulting in intelligent gadgets used in a variety of sectors, such as nursing and logistics.

REFERENCES

[1] Deep learning in multi-object detection and tracking: state of the art Sankar K. Pal1 · Anima Pramanik2 · J. Maiti2 · Pabitra Mitra3
[2] Amit Y (2002) 2D object detection and recognition: models, algorithms and networks. MIT Press, Cambridge.
[3] Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z, Qu R (2019) A survey of deep learning-based object detection. IEEE Access 7:128837–128868
[4] Pal S K (2018) Data science and technology: challenges, opportunities and national relevance. 14th annual convocation speech, national institute of technology, Calicut
[5] Chakraborty DB, Pal S K (2021) Granular Video Computing: with Rough Sets, Deep Learning and in IoT. World Scientific, Singapore
[6] Liu Y, Cheng M-M, Hu X, Wang K, Bai X (2017) Richer convolutional features for edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3000–3009
[7] Deravi F, Pal S K (1983) Grey level thresholding using second-order statistics. Pattern Recogn Lett 1(5-6):417–422
[8] Masi I, Wu Y, Hassner T, Natarajan P (2018) Deep face recognition: A survey. In: 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, pp 471–478
[9] Brunetti A, Buongiorno D, Trotta G F, Bevilacqua V (2018) Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing 300:17–33
[10] Pal N R, Pal S K (1993) A review on image segmentation techniques. Pattern Recogn 26 (9):1277–1294
[11] Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 3354–3361
[12] Krizhevsky A, Sutskever I, Hinton G E (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
[13] Chung D, Tahboub K, Delp E J (2017) A two stream siamese convolutional neural network for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1983–1991
[14] Geng H-, Zhang H, Xue Y-, Zhou M, Xu G-, Gao Z (2017) Semantic image segmentation with fused cnn features. Optoelectron Lett 13(5):381–385
[15] Li P, Wang D, Wang L, Lu H (2018) Deep visual tracking: Review and experimental comparison. Pattern Recogn 76:323–338



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY