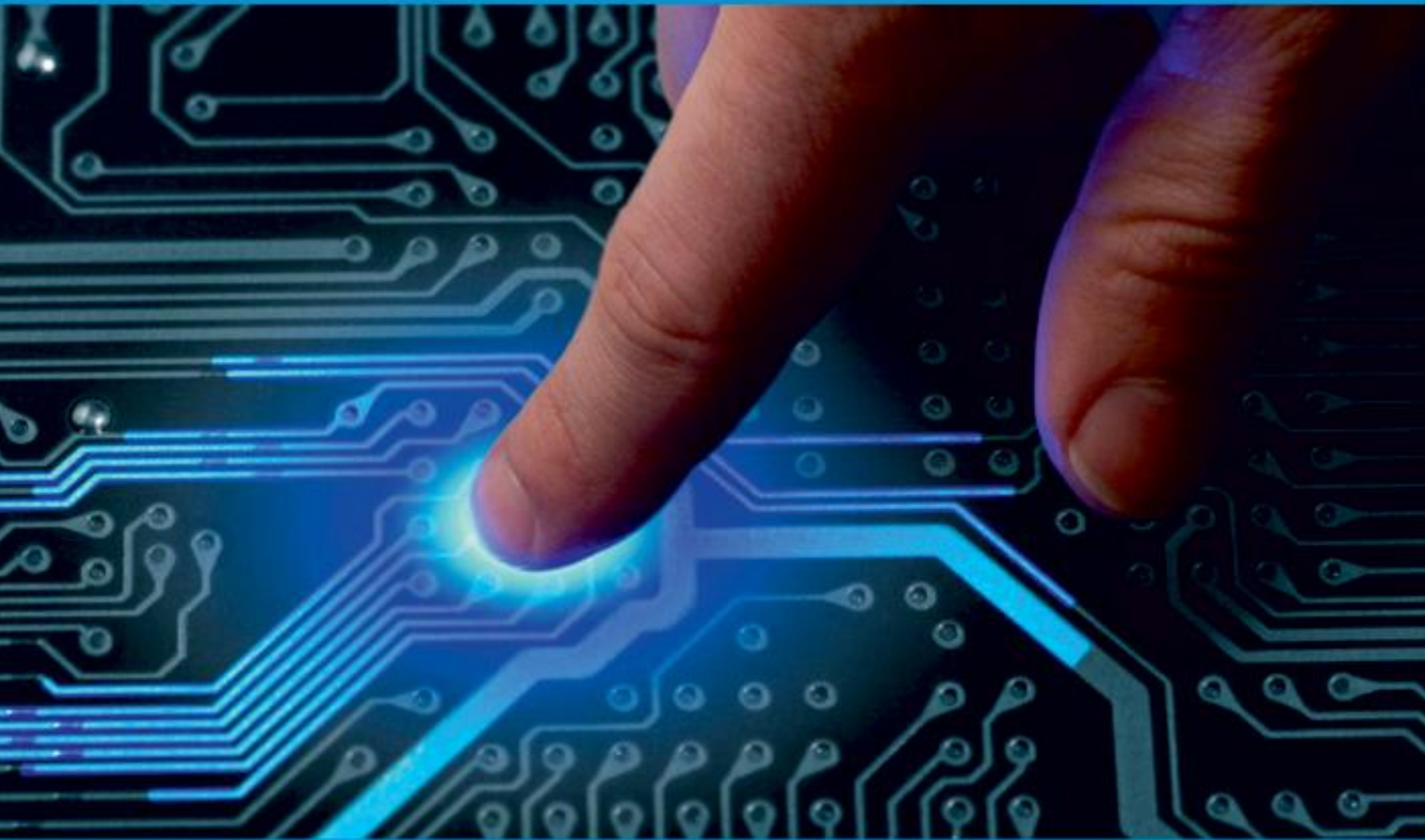
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

Real-Time Deep Learning-Based Object Enumeration and Classification: A ComprehensiveMethod for Object Detection with YOLO

Prof. S. N. Bhadane, Yash Bhadane, Mohit Chavan, Tanmay Chavan, Nachiket Pisal

Professor, Head of Department, Department of I.T., PVG's College of Engineering, Nashik, MH, India

B.E. Student, Department of Information Technology, PVG's College of Engineering, Nashik, MH, India

B.E. Student, Department of Information Technology, PVG's College of Engineering, Nashik, MH, India

B.E. Student, Department of Information Technology, PVG's College of Engineering, Nashik, MH, India

B.E. Student, Department of Information Technology, PVG's College of Engineering, Nashik, MH, India

ABSTRACT: We aim to streamline visual counting through the utilization of a real-time camera. Furthermore, we introduce a method to tally objects independently of their previous placements. All objects, including those that are immovable or intangible, can be enumerated based on this counting principle. The predominant algorithmic approach globally is the divide-and-count strategy, eliminating the need for novel insights into global image enumeration. In the past, the conventional system was the sole method for object counting; however, our proposed system accommodates varied object specifications. Each partition consists of a set of proposed regions or fixed grid cells.

KEYWORDS: Generic Method, Specific Method, YOLO Algorithm, CNN Algorithm, Images, Objects

I. INTRODUCTION

The aim of this research is to employ a real-time camera for the purpose of enumerating and distinguishing objects. A pivotal step in addressing the object counting challenge involves the object detection process, typically executed using dual flat images [1]. Enumerating objects in pictures is vital for managing traffic flow, wildlife, and human crowds. Instead of devising a specialized algorithm for each object type, we present an all-encompassing deep learning approach to object recognition. Recognizing and tallying objects within images is a complex task. Given the intricacy of this challenge, contemporary counting methods might necessitate additional annotations to accurately count the total items. The universal object counting challenge is indeed formidable [2]. The intricacy of this task has spurred the evolution of modern counting methodologies that incorporate supplementary annotations to ascertain the total count of objects. Utilizing bounding boxes, a neural network anticipates the existence of objects in an image and employs objectdetection—a more advanced version of image classification—to categorize them. Thus, "object detection" pertains to the process of pinpointing and classifying objects within an image based on predefined categories [3]. Traditional object recognition methods utilized classifiers for object detection; however, YOLO advocates for an end-to-end neural network approach that concurrently forecasts bounding boxes and class probabilities, hence the term "object detection." YOLO delivers state-of-the-art results, surpassing the performance of prior real-time object recognition systems by adopting a fundamentally distinct methodology.

II. LITERATURE DURVEY

Generic Method

As personalized treatments become more common, translational research is increasingly reliant on collaborative research efforts across multiple centres to gather enough information and resources. Counting techniques that focus on specific, interesting categories are becoming less popular than more advanced methods [4]. These contemporary techniques, unlike traditional item-counting methods, ensure the proposed model's applicability even with limited datasets categorized by type, such as those for humans, cells, or vehicles. Our goal is to count generic objects, not thesecategorized ones. This broader object-counting task has been explored in prior studies across various object types. Similar to our approach, these studies evaluate the effectiveness of counting using generic object datasets. While both methods utilize bounding boxes, we also aim to learn general item counts

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

without any additional local guidance.

Generic types refer to those with parameters. This research aims to develop categories that can interact with diverse data types. The method combines a learning layer that anticipates item counts across the entire image with the requirement for consistent counting while handling overlapping image segments [5]. When presented with an image, we categorize it into distinct sections. These sections are organized in a hierarchy, with increased depth corresponding to a more detailed examination of image areas. Our proposed method utilizes generics for counting larger, immobile objects that cannot be manipulated or moved. This straightforward approach would involve a live camera scanning an environment, analysing the objects it detects, and then classifying them based on their quantity. To assess a reading room's condition using a general approach, the entire room would first be scanned. Following analysis, the qualities and number of each object would be presented, such as the number of chairs, tables, and students.

The proposed method prioritizes scalability, aiming to function effectively across diverse environments with varying object types and densities. This adaptability ensures its applicability in real-world scenarios beyond the initial reading room example.

This research paves the way for future advancements in image analysis. While the current focus is on object quantification, the framework could be extended to include additional functionalities. It could potentially identify object types or even assess their condition, providing a more comprehensive understanding of the visual scene.\

Specific Method

Counting individuals is a common task. Specific indicators related to individuals are often utilized in applications like sensors for detecting human head and shoulder movements, segmentation of human figures against backgrounds, and the classification of foreground and background in camera systems, among others. Some person segmentation techniques leverage video motion data. For example, recent studies incorporate advanced convolutional neural networks. Rather than relying on person-specific indicators as these systems do, we explore a group-oriented counting approach and evaluate our proposed method using general object datasets [6]. We introduce a novel positional addition technique that establishes a structured hierarchy for image segmentation. In this approach, we identify visually distinct regions representing objects. By implementing a hierarchical image segmentation framework, we capitalize on the multi-level granularity inherent in our concept. An unsupervised proposal method provides a promising starting point due to its inherent hierarchical nature. Multiple proposal techniques are available for consideration [7]. The second approach involves creating a region hierarchy through incremental addition. In this context, our regions act as proposals for individual objects. Thus, our image segmentation hierarchy is anchored on two principles: (I) each segment should fully encompass the image, and (II) segments from one division may partially overlap with segments from another division without complete overlap. The first principle necessitates the use of a universal counting annotation [8]. The second allows for segmenting the image into increasingly detailed granularities. The hierarchical segmentation process concludes when no further segments can be added without being fully contained within another segment. Both the number of segments per image and the granularity of each segment hierarchy can be tailored. Previously, the only method available involved scanning every image captured by the camera. To specifically identify and tally certain objects, we have devised a specialized technique for our proposed system. Consider a classroom as an example. If our objective was solely to count computers, the proposed method would only register and document the visible computer screens. While this approach was effective for capturing specific details in crowded environments, such as densely packed lectures, it could also be employed to identify a particular item within a crowd.

III. OBJECTIVE STATEMENTS

- The aim of the project is to develop a platform where users can upload images and analyse their characteristics.
- To locate the object and ascertain its quantity.
- Design a model that has been validated with diverse image datasets to swiftly recognize and enumerate objects withinan image.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

IV. ALGORITHM

YOLO Algorithm

Utilizing neural networks, the YOLO technique facilitates real-time object identification. The algorithm's renown stems from its precision and speed. It has been deployed across various applications to differentiate between animals, humans, parking meters, and traffic signals [9]. The acronym YOLO stands for You Only Look Once. This software analyses and identifies multiple elements within an image instantaneously. In YOLO, object recognition is executed asa regression task, offering probabilities for each identified object class. YOLO leverages convolutional neural networks(CNN) for rapid object detection. True to its name, the approach necessitates only a single pass through the neural network to identify objects. This implies that a single algorithmic execution can manage predictions for the entire image. The CNN is utilized to simultaneously predict multiple class probabilities and bounding boxes.

Several iterations of the YOLO algorithm exist, with Tiny YOLO and YOLOv3 being particularly popular. A neural network anticipates and identifies objects within an image using bounding boxes, a sophisticated type of image categorization. Hence, "object detection" pertains to pinpointing and delineating objects in an image based on multiple predefined categories. Object detection, also known as object recognition, holds significant importance in computer vision due to its myriad real-world applications, including detection, recognition, and localization tasks [10]. Typically, object detection finds application in autonomous driving systems. In such systems, a combination of LIDAR, computer vision, and other technologies is employed to generate a three-dimensional map of the environment, encompassing all its entities. Moreover, it is frequently deployed in video surveillance, especially for monitoring crowds to mitigate potential security threats, tallying individuals for demographic studies, or enhancing user navigation within commercial complexes. Image classification comprises three progressively complex stages: image categorization, object localization, and object detection. YOLO [11] segments the image into N grids, each mandated to encompass a consistent S*S area. These N grids are responsible for detecting and localizing the objects they encompass. It operates on a regression principle, predicting both categories and bounding boxes for the entire image in a singular algorithmic iteration, rather than focusing solely on significant regions. The prevailing object detection method, namely YOLO, is both rapid and precise. Imagine procuring bulk materials at a construction site; manually tallying each item is daunting [12]. The YOLO Algorithm streamlines this process by identifying and cataloguing each component just once. Subsequently, a detailed count and specification analysis can be conducted.

V. SYSTEM DESIGN

The previous chapter outlines the project's inception based on an extensive review of existing literature and the methodologies planned for its execution. The summary offers a brief overview of the studies that were evaluated and examined, whereas the review offers a comprehensive perspective on the subject matter.

System Architecture

The primary process is initiated by the user, as they are the ones requiring the object enumeration. Live cameras capture the input, which is then processed based on the user's specifications. The dataset progresses through the subsequent phases:

Data Cleansing involves eliminating redundant information from the original document to expedite the processing of the retained data.

Feature Generation: The initial data is converted into numerical attributes suitable for subsequent processing stages. Image Partitioning: A crucial step in photo processing involves breaking down larger images into smaller segments tosimplify their complexity.

The YOLO (You Only Look Once) Algorithm is utilized for object localization. Identifying objects within images is atypical function of computer vision applications.

Once the object is detected and enumerated using the above-mentioned methods, the user is presented with theoutcomes.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

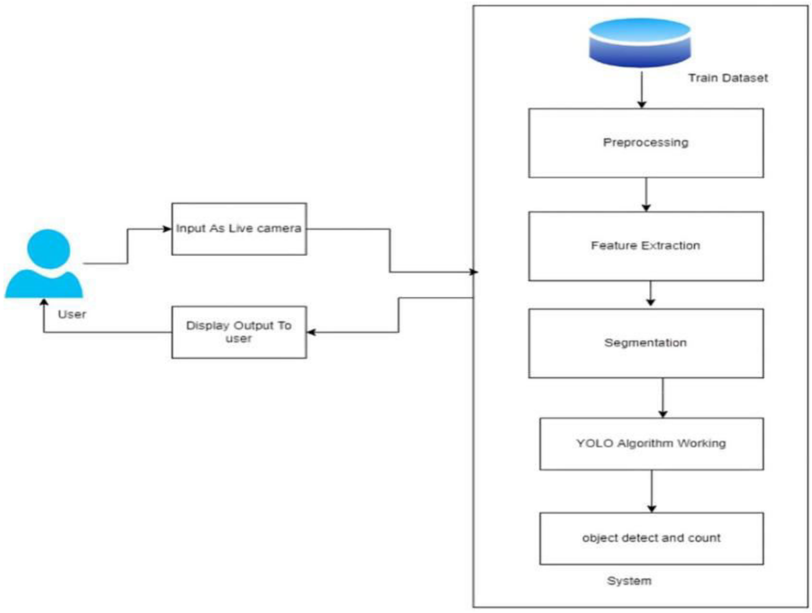


Fig.1. System Architecture

VI. RESULTS

Home page: The user has access to the options to Register, Login, and Exit on this page. The user can communicatewith the system's main operational module using these settings.

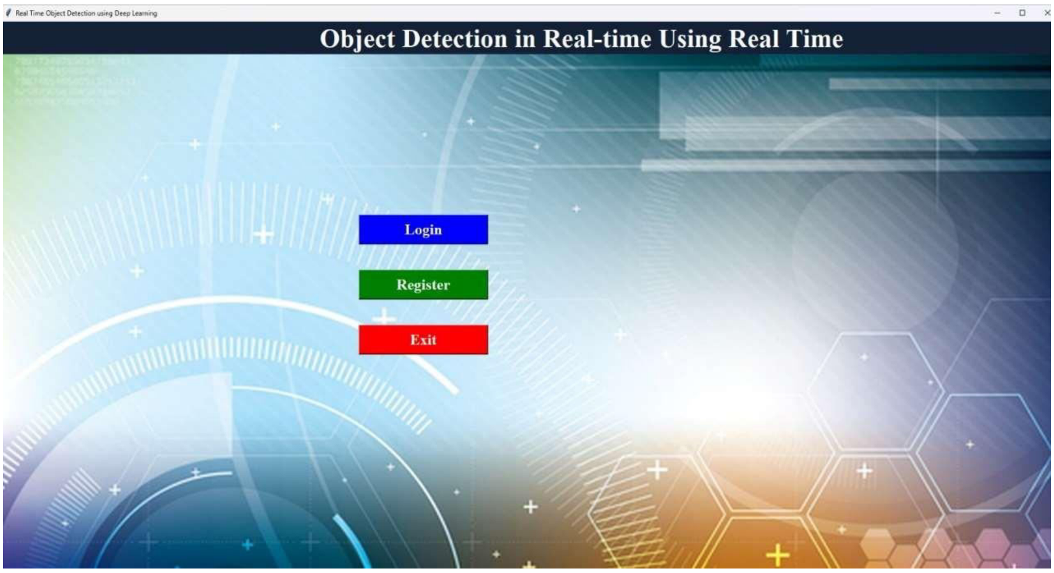


Fig.2. Home Page

Registration Page: The user provides their registration details on this page, which are utilized to enrol them in thesystem.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

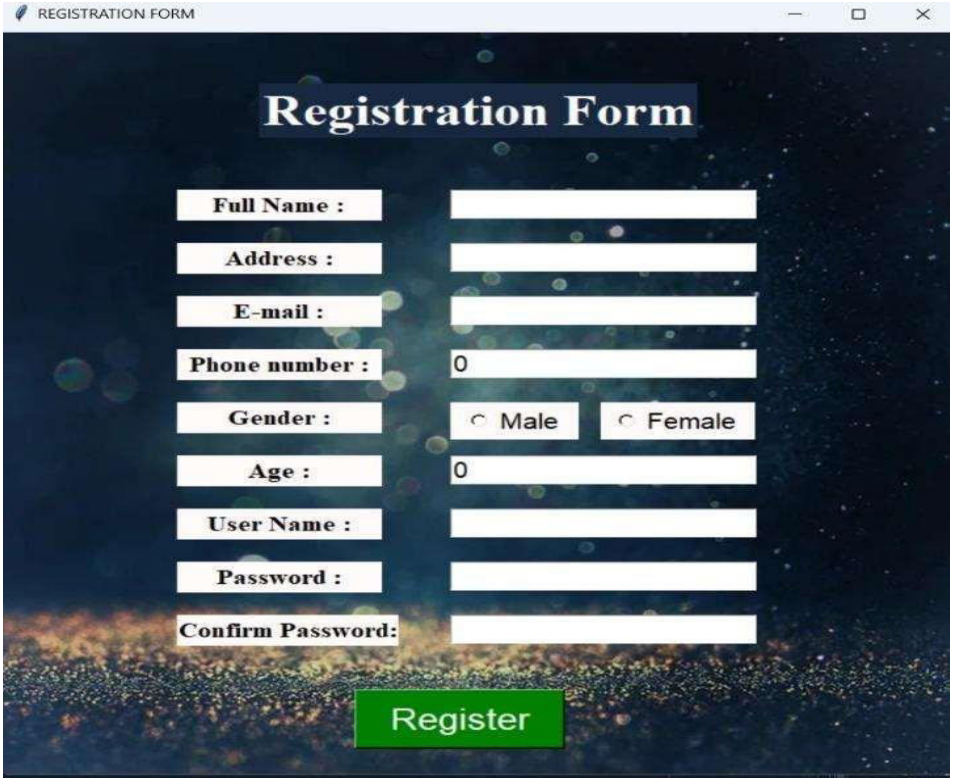


Fig.2. Registration Page

Login Page: The user submitted their credentials on this page to access their account. Only users with valid logindetails can access the system's core features.

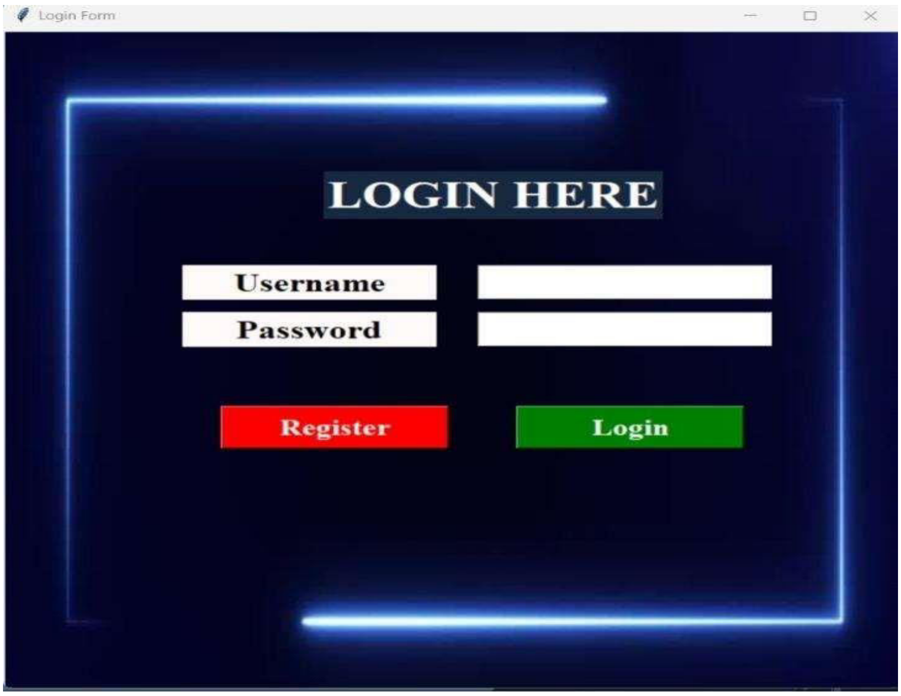


Fig.3. Login Page

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

Dashboard: After logging into the system using approved credentials, the user encounters this screen, which providesthree options.

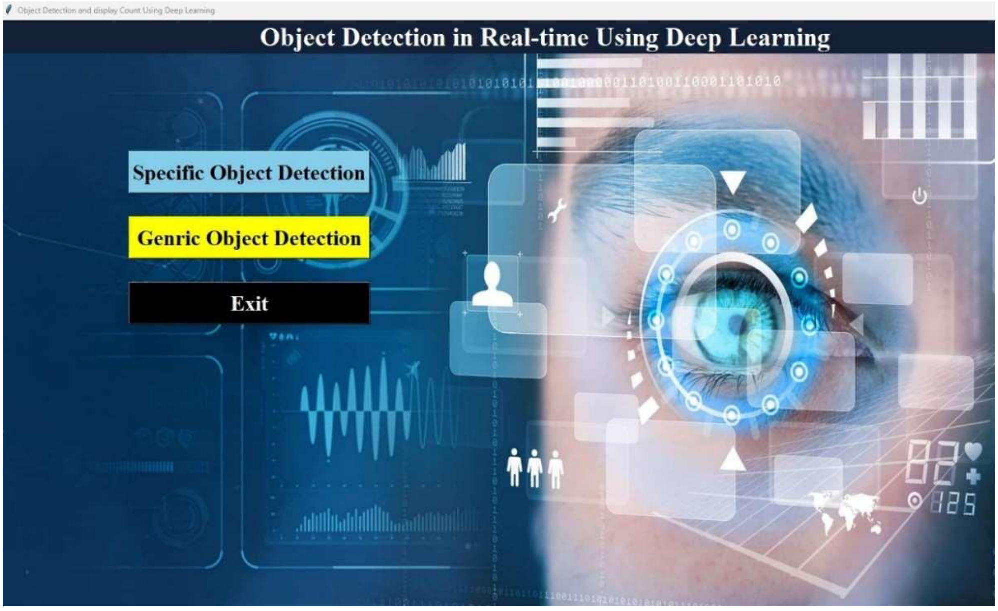


Fig.4. Dashboard

In Real Life Outputs:

Fig.5. Identifying Multiple Objects

Fig.6. Identifying Type of Vehicle

1. **General Object Recognition**: Upon choosing this option, the camera activates and starts assessing the objects within its view. It identifies the name, quantity, and distance of the objects from the camera. This feature is designed to recognize commonly encountered items.
2. **Targeted Object Identification**: This feature is utilized to pinpoint a particular object. It is employed for specialized tasks. The option displays the name, quantity, and proximity of the identified objects. However, this function is specifically trained for distinct items.

VII. CONCLUSION

This research proposes a method for object detection and enumeration using a real-time webcam. To integrate geometric data into the loss function, we utilize object proposals or regular grids along with the principle of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205252 |

inclusion- exclusion. Additionally, we advocate for both local learning from image features and global learning from the completepicture.

REFERENCES

1. C. Arteta, V. Lempitsky, and A. Zisserman. Counting in the wild. In ECCV, pages 483–498, 2016.
2. A. B. Chan, Z. S. J. Liang, and N. Vasconcelos. Privacy-preserving crowd monitoring: Counting people without people models or tracking. In CVPR, pages 1–7, 2008.
3. A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. PAMI, 30(5):909–926, 2008.
4. A. B. Chan and N. Vasconcelos. Bayesian poisson regression for crowd counting. In ICCV, pages 545–551, 2009.
5. S. Chen, A. Fern, and S. Todorovic. Person count localization in videos from noisy foreground and detections. In CVPR, pages 1364–1372, 2015.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING