Volume 11, Issue 2, February – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

An Intelligent Assistive and Monitoring System for Elderly and Visually Impaired Users

Thanh Tuan Ton That[1]; Khang Huy Ngo[2]; Nam Hien Cao Nguyen[3]

[1,2,3]Phu Nhuan High School, Ho Chi Minh City, Vietnam

Publication Date: 2026/02/25

Abstract: Visual impairment and age-related decline significantly affect independent mobility and safety in daily life. This paper presents a smart monitoring and assistance system aimed at supporting visually impaired individuals and elderly people through the integrating of artificial intelligence, computer vision, and sensor-based technologies. Getting inspired by the working mechanism of the human eye, the proposed system employs the YOLO11 deep learning model for real-time object detection and classification, combined with the Depth Anything v2 model for monocular depth estimation to calculate the distance between users and surrounding objects.

The system is implemented using an embedded camera and IoT-based sensors, including ultrasonic distance sensors, GNSS positioning, heart rate monitoring (MAX30102), and fall detection modules, enabling comprehensive environmental perception and user health monitoring. Experimental evaluations were conducted in both bright and low-light environments using a self-collected dataset. The results depict that the proposed system achieves an overall the accurate of detection – approximately 95 percent, with stable performance across varying lighting conditions.

The findings confirm the feasibility and effectiveness of integrating deep learning models with embedded hardware to provide real-time assistance. This system has strong potential for development into a wearable smart device capable of enhancing mobility, reducing collision risks, and improving the independence and quality of life for visually impaired and elderly users. Moreover, the proposed approach contributes to the advancement of human-centered intelligent assistive technologies with meaningful social impact.

Keywords: *Assistive Technology, Visual Impairment, Elderly Care, Object Detection, Depth Estimation, YOLO, Computer Vision, Internet of Things (IoT).*

How to Cite: Thanh Tuan Ton That; Khang Huy Ngo; Nam Hien Cao Nguyen (2026) An Intelligent Assistive and Monitoring System for Elderly and Visually Impaired Users. *International Journal of Innovative Science and Research Technology*, 11(2), 1438-1459. https://doi.org/10.38124/ijisrt/26feb637

I. INTRODUCTION

Vision plays a fundamental and irreplaceable role in human life, enabling individuals to perceive their surroundings, interpret information, and navigate safely in daily activities. However, visual impairment and age-related physical decline pose significant challenges to independent mobility, personal safety, and overall quality of life. According to global statistics, approximately 43 million people [2] are blind and more than 295 million suffer from moderate to severe visual impairment worldwide [2]. In Vietnam alone, there are around 2 million people with serious visual impairment and over 12.5 million elderly individuals aged 60 and above. These numbers highlight an urgent need for effective assistive solutions that support vulnerable populations in maintaining autonomy and safety. [2]

With the rapid advancement of artificial intelligence and embedded systems, assistive technologies have gained increasing attention as a means to bridge the gap between human limitations and environmental demands. Although several support tools currently exist—such as mobilebased collision detection systems—most of these solutions are limited in functionality and are not designed based on the biological working principles of human vision. Furthermore, many vision-related technologies today primarily serve immersive or entertainment purposes, such as virtual and augmented reality devices, rather than addressing real-world safety and mobility challenges faced by visually impaired individuals and the elderly. [6]

Getting motivation by these limitations, this study proposes an Intelligent Assistive and Monitoring System for

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

Elderly and Visually Impaired Users, inspired by the functional mechanisms of the human visual system. The proposed system integrates camera-based perception, deep learning algorithms, and sensor fusion to observe the surrounding environment, recognize objects, estimate distances, and provide timely warnings of potential hazards. In addition to environmental perception, the system is designed to support health monitoring features such as fall detection, heart rate measurement, and location tracking, thereby enhancing user safety and enabling rapid emergency response. [1]

Despite its potential, the development of such a system presents several technical challenges. The performance of visual monitoring systems is highly influenced by external factors including lighting conditions, camera resolution, user movement, and environmental complexity. While high-quality cameras can improve image acquisition, the overall effectiveness of the system remains strongly dependent on robust algorithms, adaptable models, and realistic usage scenarios. Addressing these challenges is therefore a critical component of this research.

Through theoretical analysis, system design, experimental implementation, and performance evaluation, this work aims to demonstrate the feasibility and practicality of an integrated assistive solution. The proposed approach not only contributes to the development of intelligent humancentered technologies but also holds significant social value by supporting independent living and improving the quality of life for visually impaired individuals and elderly users.

II. CHAPTER 1: THEORETICAL FOUNDATION AND RELATED STUDIES

This chapter provides a comprehensive synthesis of the current research landscape and the fundamental theories governing optical acquisition and computer vision. By establishing a robust theoretical framework, this section serves as the prerequisite for developing intelligent sensorbased solutions aimed at assisting the elderly and individuals with visual impairments.

➢ *Fundamentals of Camera Operation and Modeling*

We examine the technical principles of image acquisition, focusing on how camera parameters—such as focal length, sensor sensitivity, and field of view—impact the reliability of data. Understanding these mechanics is vital for ensuring that the system can operate effectively under varying environmental constraints and lighting conditions.

➢ *Human Behavioral and Feature Analysis*

A critical component of this research involves identifying and modeling patterns in human characteristics. This includes the study of physical traits, postures, and movement trajectories. Analyzing these features allows the system to interpret user intent and distinguish between routine activities and potential hazards.

➢ *Image Databases and Classification Methodologies*

The efficacy of deep learning models is deeply rooted in the quality of the training data. This section explores:

- *Dataset Integration:*

The utilization of comprehensive image repositories containing diverse samples of human subjects and everyday objects.

- *Entity Classification:*

The implementation of algorithms designed to distinguish between static objects (fixed obstacles) and dynamic objects (moving entities such as pedestrians or vehicles).

➢ *Prerequisites for Assistive System Design*

The theoretical concepts presented in this chapter are not merely academic; they form the functional foundation for the proposed assistive device. By synthesizing these principles, the system aims to provide a reliable ”digital perception” layer, significantly enhancing the mobility, safety, and independence of the elderly and the visually impaired.

➢ *Image Databases and Methodology*

The effectiveness of any computer vision system is rooted in the quality of its training data. This study utilizes diverse image repositories containing samples of everyday objects and human subjects.

- *Object Classification:*

A core focus is placed on the methodologies used to distinguish between static objects (such as furniture or walls) and dynamic objects (such as pedestrians, cyclists, or vehicles).

- *Experimental Process:*

We detail the pipeline from raw data acquisition to pre-processing, feature trilateralization, and final classification.

➢ *Foundations for Assistive Application*

The theoretical concepts presented here serve as the prerequisite for the practical development of a smart sensor device. By integrating these basic principles, the system is designed to act as an ”artificial eye,” enhancing the mobility, reducing collision risks, and improving the overall independence of individuals with visual impairments and the elderly.

III. METHODOLOGY: CONVOLUTIONAL NEURAL NETWORKS (CNN)

A Convolutional Neural Network (CNN), or ConvNet, is a specialized deep learning architecture designed for the automated extraction of spatial hierarchies from high-dimensional data. CNNs have become the gold standard for computer vision tasks, particularly in image classification and object recognition, due to their ability to preserve local dependencies and spatial topology within an image. [12]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

➢ *Data Representation*

From a computational perspective, an input image is treated as a three-dimensional tensor defined by the parameters ($W \times H \times D$), where:

- W (Width): The horizontal pixel resolution.
- H (Height): The vertical pixel resolution.
- D (Depth): The number of color channels (e.g., $D = 3$ for RGB or $D = 1$ for grayscale).

The network interprets these dimensions as a matrix of numerical pixel intensities, which serve as the raw input for the feature learning process.

➢ *Architectural Components*

The architecture of a CNN is traditionally organized into a sequence of functional layers that transform raw pixels into high-level semantic features:

- *Convolutional Layer:*

This layer serves as the primary feature extractor. It utilizes a set of learnable kernels (filters) that perform element-wise multiplication and summation—a convolution operation—across the input to produce feature maps.

- *Pooling Layer:*

To reduce the spatial dimensionality of the feature maps and minimize computational overhead, pooling layers (typically Max or Average Pooling) are employed. This process also enhances the model's translational invariance.

- *Fully Connected (FC) Layer:*

Following the feature extraction stages, the flattened feature maps are passed into one or more FC layers. These layers aggregate the learned features to perform the final non-linear mapping to the output classes.

➢ *Classification and Activation*

At the final output stage, a Softmax activation function is applied to the raw logit scores. The Softmax function normalizes these scores into a categorical probability distribution, as defined in Equation 2:

$$P(y = i \mid \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

Where z_i represents the output score for class i, and K denotes the total number of target categories. This allows the model to determine the most probable class for the input image based on the extracted feature set.

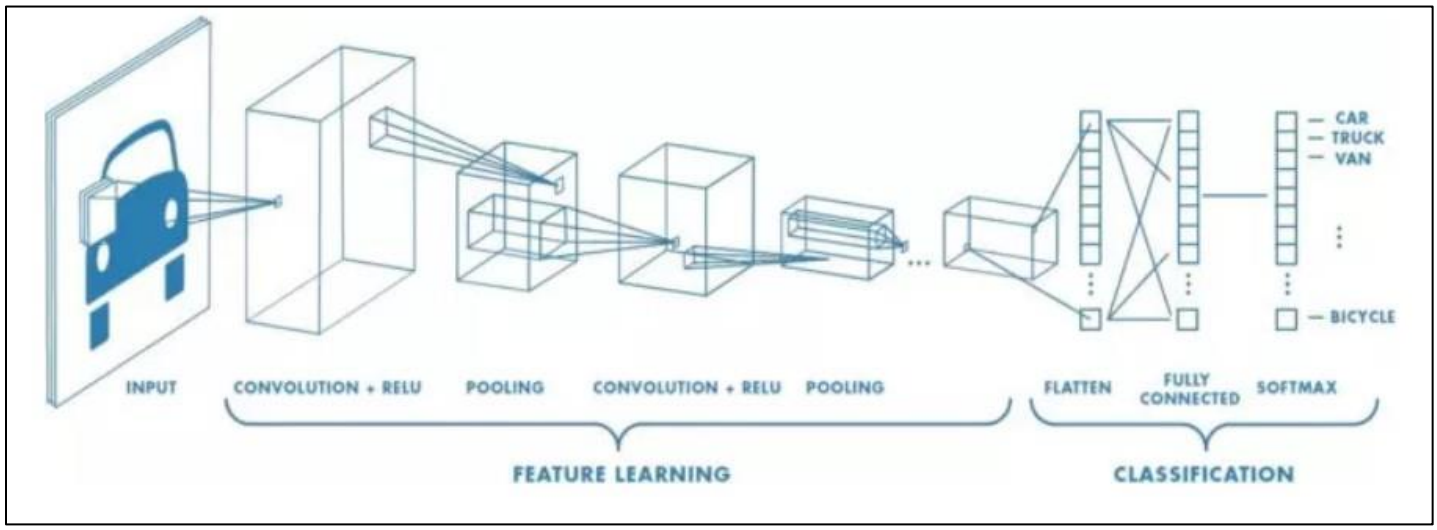

Fig 1 CNN Model Used for Problem Analysis

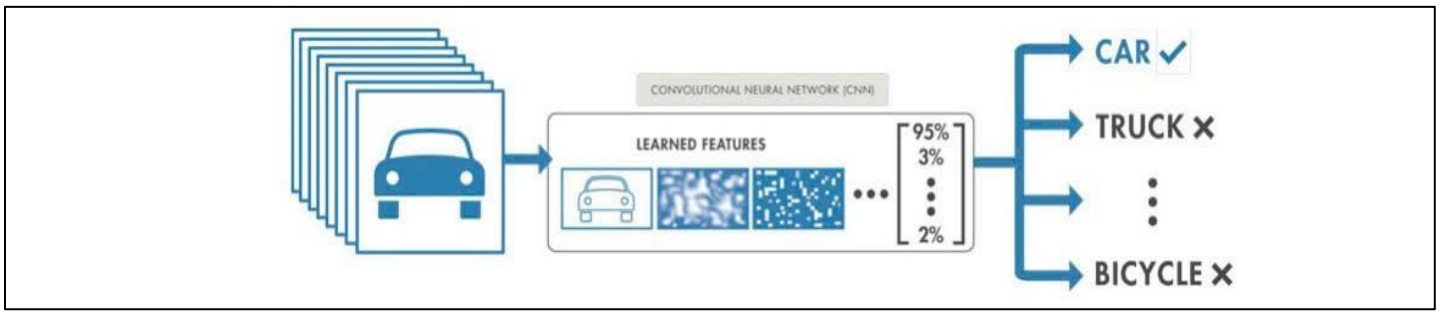

Fig 2 CNN Model Used for Problem Analysis[10]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

IV. ARCHITECTURE OF THE YOLOV11 MODEL

YOLO (You Only Look Once) is a Convolutional Neural Network (CNN)-based object detection model designed for real-time object recognition. It can also be considered an evolution of RegionBased Convolutional Neural Networks (R-CNN) in addressing object localization and detection tasks [7].

YOLO is regarded as a simple yet powerful algorithm. It formulates object detection as a single regression problem applied to the entire image. This means that both object locations and their corresponding class labels are predicted directly by passing the image pixel values through a single neural network.

As previously mentioned, YOLO is a CNN-based model for object detection and recognition, as it combines convolutional layers and fully connected layers. The convolutional layers extract image features, while the fully connected layers predict class probabilities and bounding box coordinates. [8]

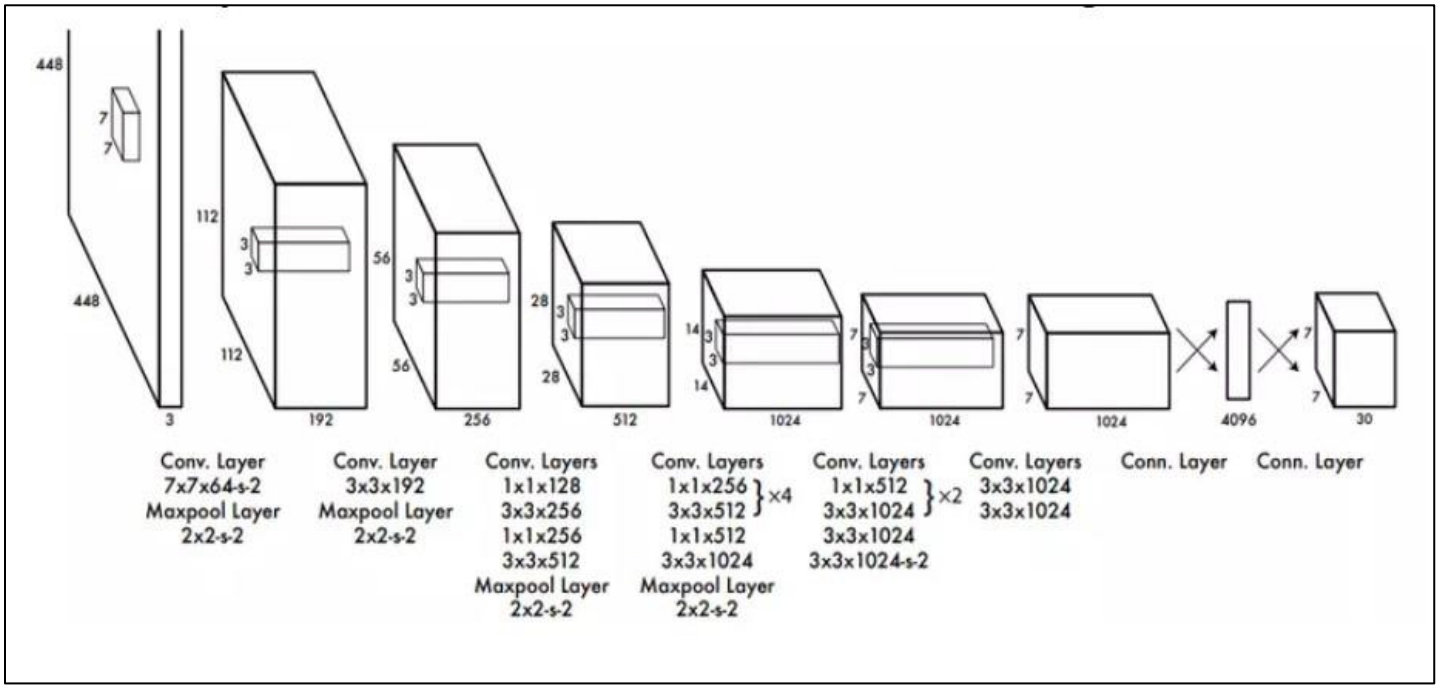

Fig 3 The Operating Process of the YOLO Network

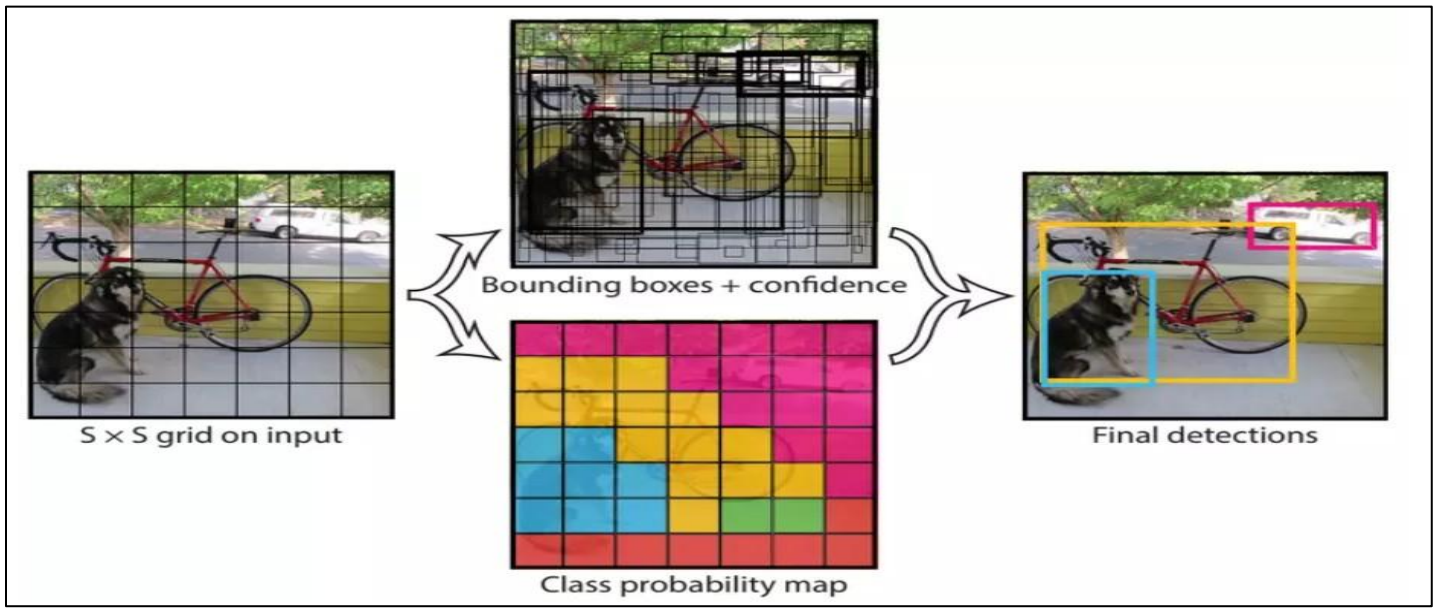

Fig 4 The Operating Process of the YOLO Network

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

V. HOW YOLOV11 WORKS

This section presents the theoretical foundation underlying YOLOv11, beginning with the general principles of Convolutional Neural Networks (CNNs), which serve as the backbone of modern object detection systems. [13]

➢ *Convolutional Neural Networks (CNN)*

A Convolutional Neural Network (CNN) is a specialized deep learning architecture designed to automatically extract spatial hierarchies of features from high-dimensional input data.

From a computational perspective, an input image is represented as a three-dimensional tensor:

$$(W \times H \times D)$$

Where:

- W denotes the image width,
- H denotes the image height,
- D represents the depth (i.e., number of color channels).

A standard CNN architecture consists of three primary components:

- Convolutional Layer

Applies learnable kernels to perform feature extraction through convolution operations, producing feature maps.

- Pooling Layer

Reduces spatial dimensions, lowering computational complexity while improving translational invariance.

- Fully Connected (FC) Layer

Maps extracted features into output space through non-linear transformations.

At the output stage, the Softmax activation function converts raw network outputs into a categorical probability distribution:

$$P(y = i \mid \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (2)$$

Where:

- z_i represents the logit corresponding to class i,
- K is the total number of classes.

➢ *The YOLO (You Only Look Once) Framework*

The YOLO framework represents a paradigm shift in real-time object detection by utilizing a single-stage pipeline. Unlike Region-Based CNNs (R-CNN) [11], which rely on separate region proposal networks, YOLO formulates detection as a unified regression problem [9]. This allows the model to predict bounding box coordinates and class probabilities directly from raw pixels in a single forward pass.

➢ *Detection Mechanism and Tensor Representation*

The core mechanism of YOLO involves the spatial discretization of the input image into an $S \times S$ grid. If the centroid of an object falls within a specific grid cell, that cell is responsible for detecting the object [16]. Each cell predicts B bounding boxes, where each box is defined by a five-parameter vector:

$$V_{box} = [x,y,w,h,c] \quad (3)$$

Where:

- (x,y): Coordinates of the object's center relative to the grid cell.
- (w,h): Width and height of the box relative to the image dimensions.
- c (Confidence): Represents the probability of an object's presence and the Intersection over Union (IoU) between the prediction and ground truth.

In addition to these parameters, each cell predicts C class probabilities. The final output of the model is encoded as a multidimensional tensor of shape:

$$X \in RS \times S \times (B \times 5 + C) \quad (4)$$

This allows the system to achieve high-speed inference, making it suitable for real-time assistive hardware.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

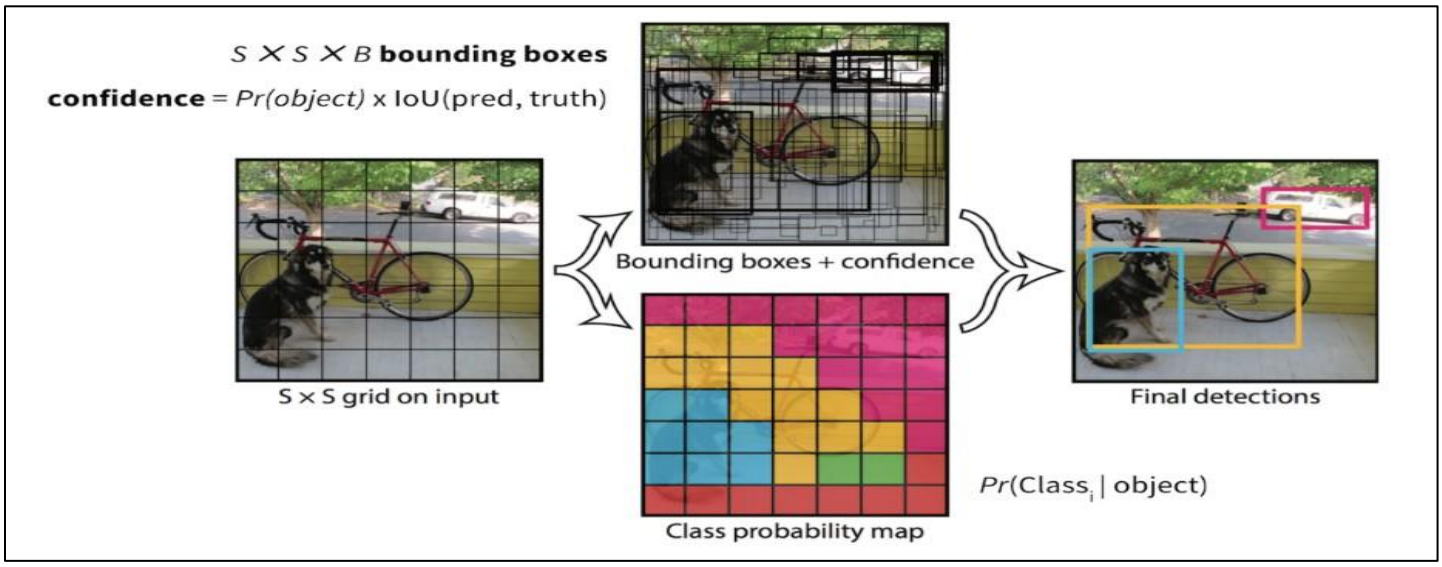

Fig 5 The Operating Process of the YOLO Network

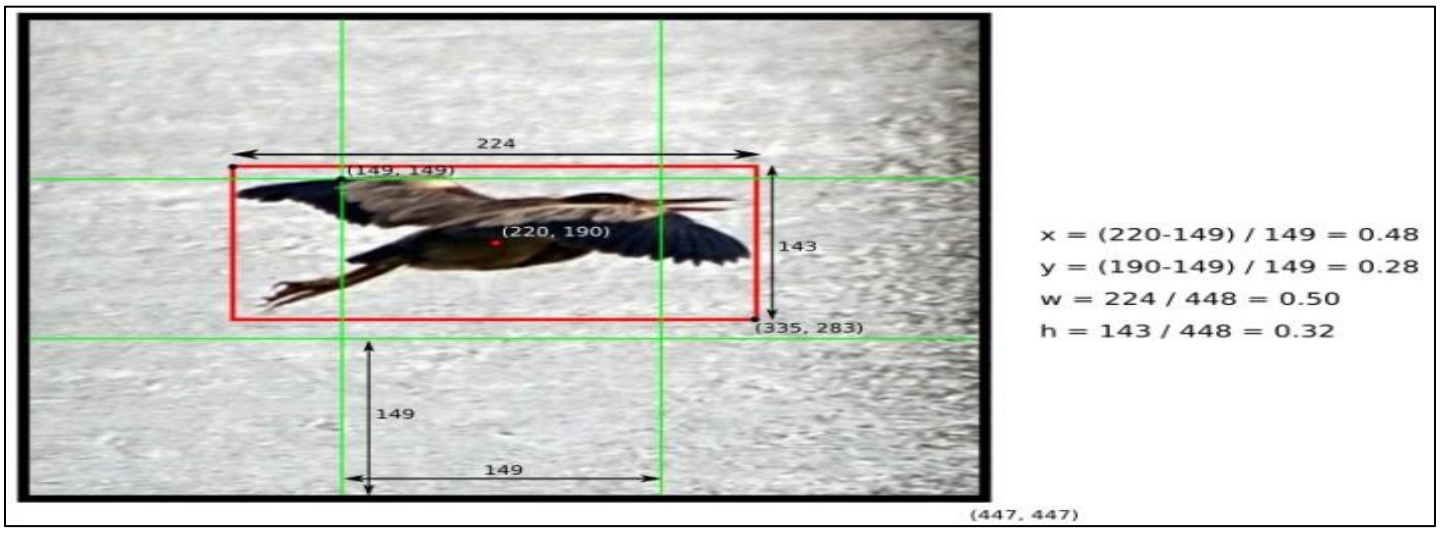

Fig 6 Parameter Values of Bounding Box in Grid

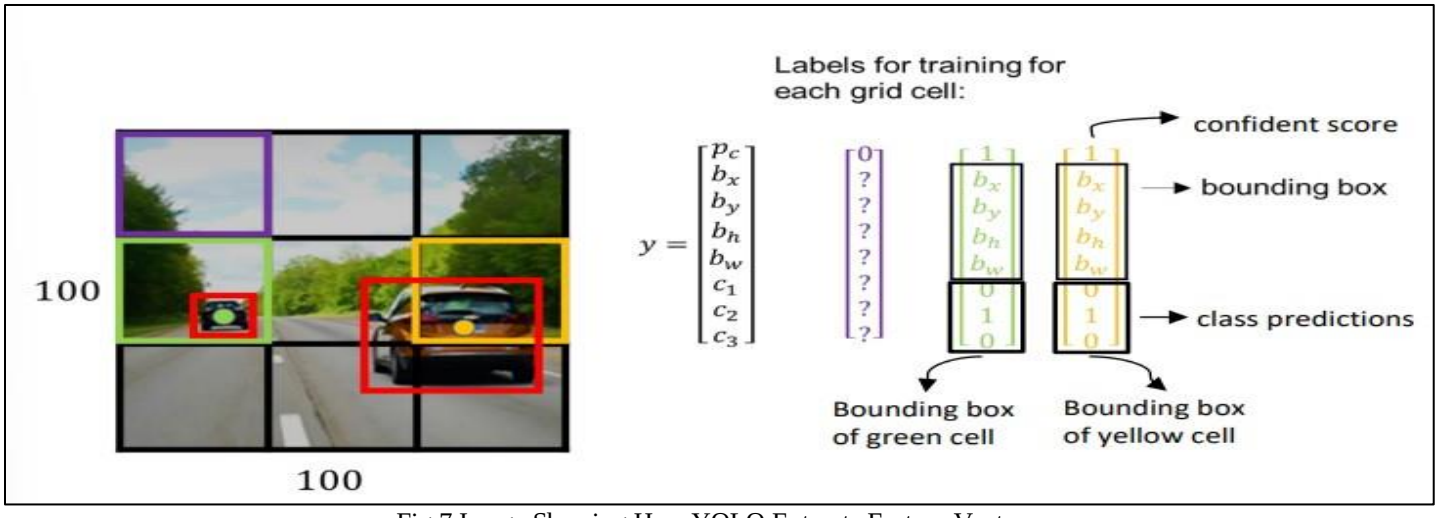

Fig 7 Image Showing How YOLO Extracts Feature Vectors.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

VI. IMAGE CLASSIFICATION CAPABILITIES USING MACHINE LEARNING AND CAMERA MODELING

According to D. Lu and Q. Weng [5], image classification using machine learning models yields highly reliable results and therefore holds significant practical potential. Below are several examples illustrating its usefulness:

- *Autonomous Vehicles:*
  Employ image classification to identify surrounding objects such as vegetation, pedestrians, traffic lights, and other environmental elements.
- *Digital Asset Management:*
  Enables users to efficiently organize and manage personal photo collections.
- *Security and Personalization:*
  Enhances security applications in smartphones and supports recommendation systems.
- *Healthcare:*
  Assists in analyzing medical images and suggesting whether they correspond to specific disease-related symptoms.

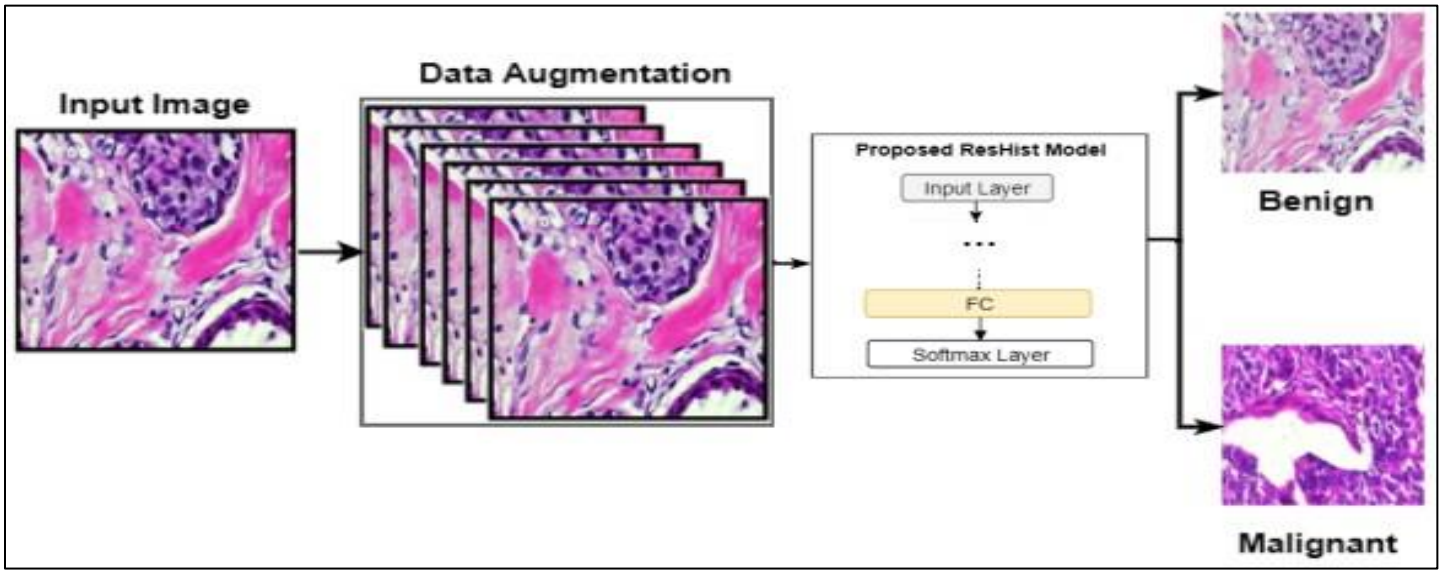

Fig 8 Classification of Breast Cancer Images Based on YOLO Theory

VII. STEPS IN IMAGE CLASSIFICATION USING DEEP LEARNING MODELS

Table 1 Workflow Stages of the Deep Learning Pipeline (Based on Figure 1.5.1)

Stage	Component	Functional Description
1	Input	Raw image data acquisition.
2	Pre-processing	Data normalization, resizing, and noise reduction.
3	Feature Extraction	Automated hierarchical learning through multiple hidden layers.
4	Recognition	Classification of objects based on learned feature vectors.
5	Output	Generation of final predicted class or probability.

Table 2 Workflow comparison: Traditional Machine Learning vs. Deep Learning.

Stage	Traditional Machine Learning	Deep Learning
1. Data Input	Input	Input
2. Processing	—	Pre-processing
3. Features	Manual Extraction	Multi-layer Extraction
4. Analysis	Prediction from Features	Recognition from Features
5. Result	Output)	

The following diagram illustrates the three key stages involved in performing image classification using a CNN-based deep learning model.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

Table 3 Architectural Comparison Between Traditional Machine Learning and Deep Learning Pipelines

[HTML]EFEFEF Pipeline Stage	Traditional ML Approach	Deep Learning Approach
Input	Raw Image Data	Raw Image Data
Preparation	—	Pre-processing (Normalization/Resizing)
Feature Handling	Manual Extraction: Expert-defined descriptors (SIFT, HOG, etc.)	Automated Learning: Hierarchical features learned through multiple layers
Decision Making	Prediction based on static features	Recognition via end-to-end training
Output	Final Classification	Final Classification

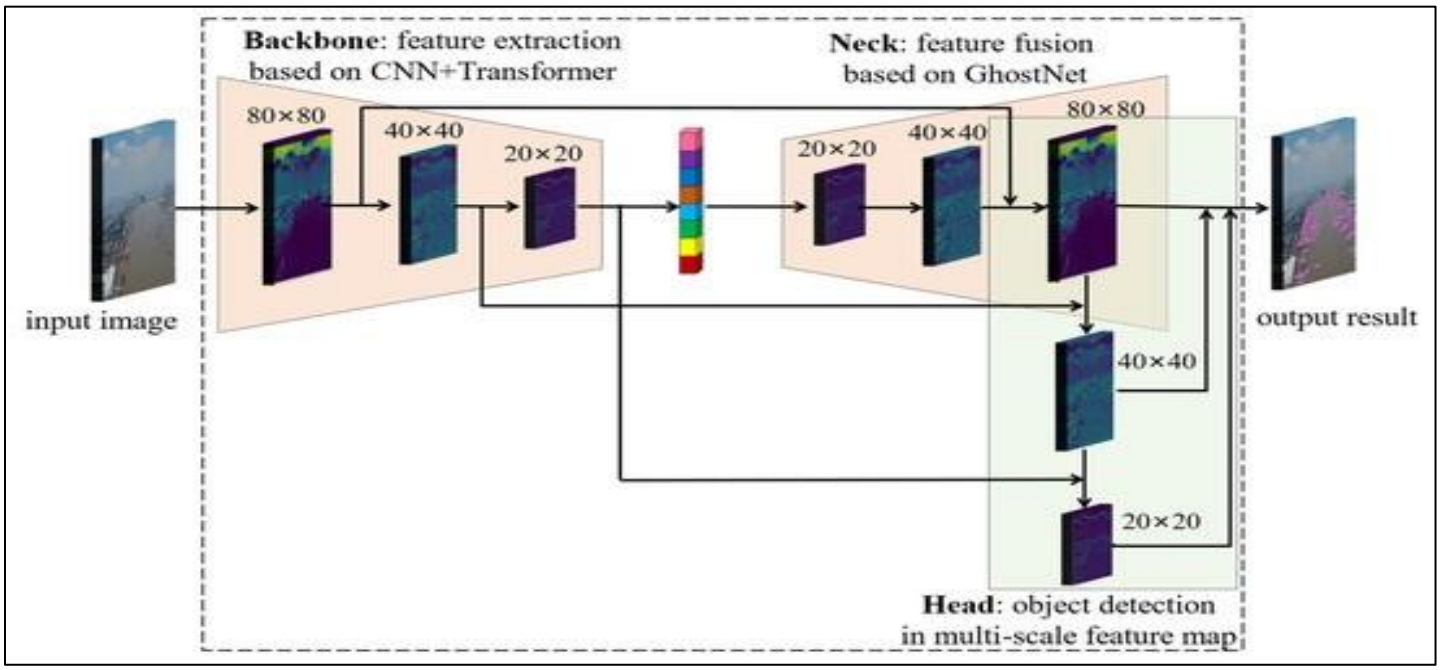

Fig 9 Image Showing How YOLO Extracts Feature Vectors. [15]

VIII. DISTANCE ESTIMATION VIA DEPTH ANYTHING V2

Depth Anything V2 represents a significant advancement in monocular depth estimation, capable of producing high-fidelity depth maps from single-frame inputs. Unlike traditional models, it demonstrates robust performance on complex surfaces, including transparent and reflective materials.

➢ *Depth Map Interpretation*

The model outputs a visual representation of spatial proximity. The mapping scheme used for relative distance is summarized in Table 4.

Table 4 Depth Map Color Encoding Scheme

Color Gradient	Relative Distance
Deep Purple	Nearest proximity
Orange-Purple	Intermediate distance
Light Yellow	Maximum distance

➢ *Architectural Framework*

The efficacy of Depth Anything V2 relies on a dual-stage pipeline:

- *Training Stage:*

Utilizes a Teacher-Student paradigm to leverage both synthetic data and unlabeled real-world imagery for superior generalization.

- *Inference Stage:*

Employs a Vision Transformer (ViT) Encoder for feature extraction and a specialized Decoder for pixel-wise depth reconstruction.

➢ *Practical Application and Calibration*

While the model excels at relative spatial ordering, absolute distance estimation requires scale calibration. For the proposed "Assistive Monitoring System," this calibration is

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

essential to translate relative values into metric distances, facilitating safer navigation for visually impaired users.

IX. DISTANCE MEASUREMENT BETWEEN OBJECTS USING LASER BEAMS

A laser is a device that stimulates atoms or molecules to emit light at specific wavelengths and amplifies that light, typically producing a highly narrow beam of radiation. This emission usually spans a limited range within the visible, infrared, or ultraviolet spectrum. The term laser stands for *Light Amplification by Stimulated Emission of Radiation*. [3]

➢ *Operating Principle*

Distance measurement using laser technology is based on the principle of Time-of-Flight (ToF). A laser beam is emitted from the source toward a target object. When the beam encounters an obstacle, it is reflected back to the sensor. The system then calculates the distance by measuring the time taken for the emitted signal to travel to the object and return.

➢ *Distance Calculation Formula*

The distance *d* is determined by the following equation:

$$d = \frac{c \times t}{2} \tag{5}$$

Where:

- Speed of Light (*c*): Light travels at a constant speed of approximately 299,792,458 m/s (approximated as 3×10^8 m/s).
- Time (*t*): The duration between the emission of the laser pulse and the reception of its reflected signal.
- Distance (*d*): The actual distance to the target, obtained by dividing the total travel distance by two.

Table 5 Summary of ToF Laser Measurement Components

Variable	Measurement	Role in Pipeline
c	Velocity	Constant physical parameter
t	Chronometry	Variable measured by the sensor
d	Length	Final output for monitoring system

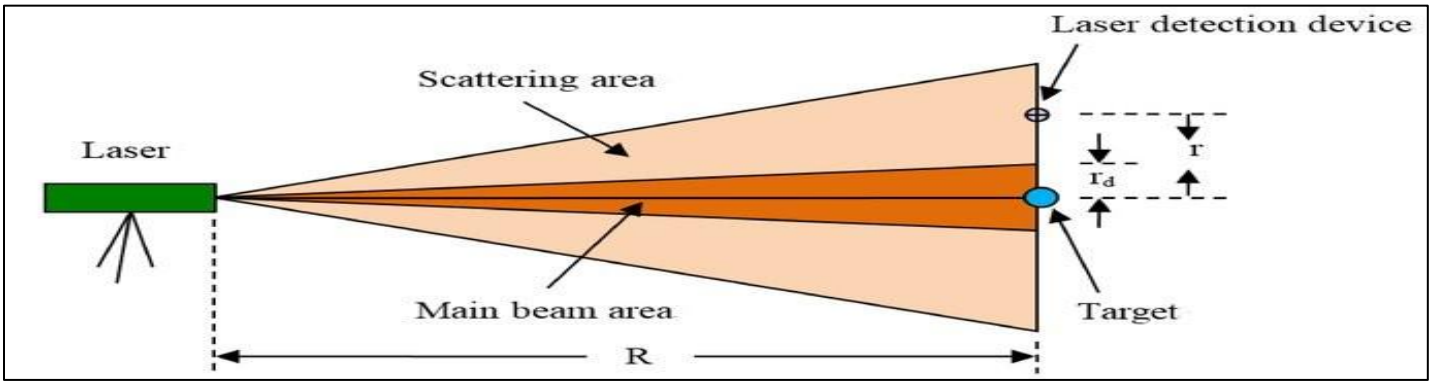

Fig 10 Simulation of Laser Distance Measurement.

Fig 11 Simulates the Measurement of the Laser Beam's Distance to the Device.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

X. MEASURING DISTANCE BETWEEN THE USER AND OBJECTS USING AN ALGORITHMIC APPROACH

The human visual system naturally perceives that objects appear larger when they are closer and smaller when they are farther away. Based on this observation, the research team derived a geometric formula to estimate distance within the system:

$$D = H \times K \quad (6)$$

Where *D* represents the perceived height of the object in the field of view (cm), *H* is the actual distance between the user and the object (cm), and *K* is a proportionality constant.

➢ *Experimental Calibration of Constant K*

The constant *K* is determined experimentally by averaging multiple measurements. Using various height-to-distance scenarios, the following values were calculated:

- $K_1 = \frac{2}{200}$(2m height at 2m distance)
- $K_2 = \frac{1.8}{400}$(1.8m height at 4m distance)
- $K_3 = \frac{1.6}{600}$(1.6m height at 6m distance)
- $K_4 = \frac{1.4}{800}$(1.4m height at 8m distance)
- $K_5 = \frac{1.2}{1000}$(1.2m height at 10m distance)
- $K_6 = \frac{1}{1200}$(1m height at 12m distance)

The resulting average constant is:

$$K = \frac{\sum_{i=1}^{6} K_i}{6} = \frac{419}{120000} \quad (7)$$

XI. CHAPTER 2. MODEL DESIGN

➢ *Design Materials*

The proposed model is fabricated using 3D-printed plastic, specifically Thermoplastic Polyurethane (TPU). This material exhibits high elasticity, corrosion resistance, thermal stability, and oil resistance, contributing to overall durability. Additionally, TPU enhances safety during experimental implementation. Its relatively low cost also helps reduce production expenses and supports the feasibility of the project. [14]

➢ *Electronic Components and Supporting Hardware*

Table 6 System Component Specifications

No.	Component	Function
1	3D Printer	Used to fabricate the physical model.
2	Breadboard	Used for prototyping electronic circuits.
3	ESP32 Camera	Serves as the vision module.
4	SX1278 LoRa RA-01	Transceiver module for distance measurement signals.
5	Laser Driver Module	Emits laser beams for distance measurement.
6	ESP32	Used for programming and system control.

➢ *Hardware Components*

The system incorporates essential electronic components, including a microcontroller for central processing, laser sensors for distance measurement, and auxiliary modules to support data acquisition and signal transmission. These components were selected based on their reliability, efficiency, and compatibility with the overall system architecture.

➢ *Artificial Intelligence Framework*

Artificial intelligence models were employed to enhance system performance and enable intelligent data processing. The trained models support accurate interpretation of sensor data and improve decision-making capability during real-time operation.

➢ *System Integration*

All hardware and software components were integrated to ensure seamless communication between sensing, processing, and response mechanisms. The integration process focused on optimizing system stability, minimizing latency, and ensuring consistent performance during testing.

➢ *Testing Environment*

Experimental validation was conducted under controlled conditions to ensure repeatability and safety. The use of simulated motion through the line-following robot allowed the system to be evaluated under dynamic scenarios resembling real-world operation.

➢ *Performance Evaluation*

System performance was assessed based on accuracy, response time, and operational stability. Multiple testing trials were conducted to validate the effectiveness of the proposed design and to identify potential areas for future improvement.

➢ *Image of Products*

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

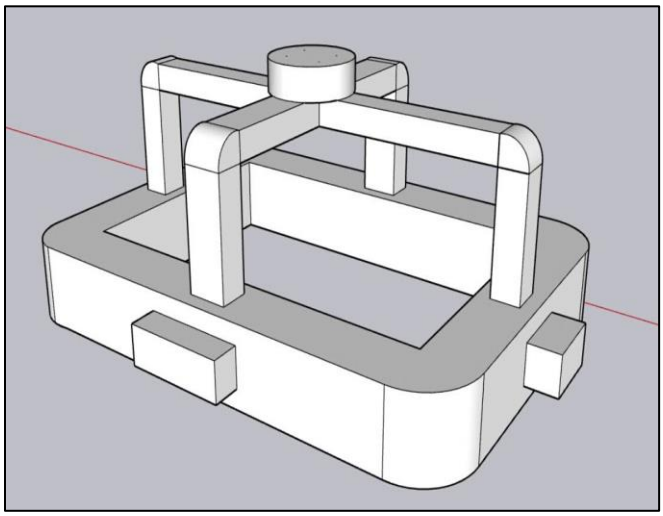

Fig 12 Design Prototype 1 Using an Acoustic Sensing System and Camera (Side View)

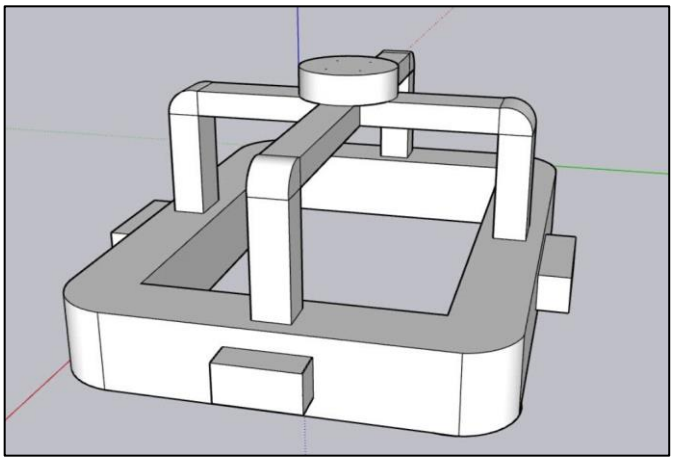

Fig 13 Design Prototype 1 Using an Acoustic Sensing System and Camera (Front View)

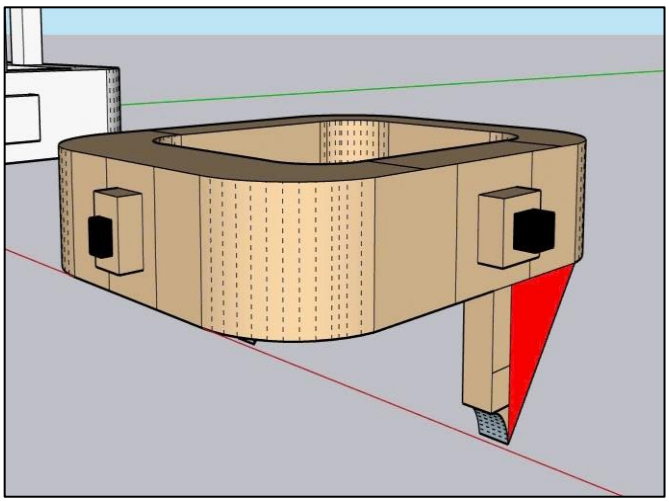

Fig 14 Design Prototype 2 Using Camera Only (Right View)

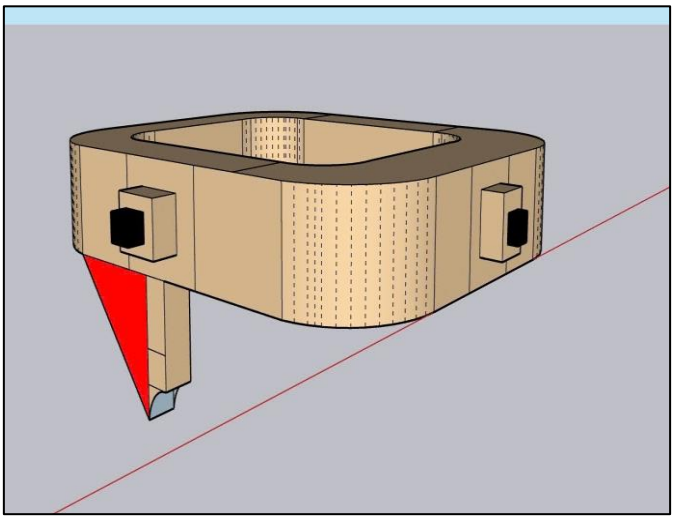

Fig 15 Design Prototype 2 Using Camera Only (Left View)

XII. CHAPTER 3. EXPERIMENTATION AND EVALUATION

➢ *Experimental Environment*

The experimentation was conducted on two distinct computing systems to verify performance across different architectures:

- Computer 1: Windows 11 (64-bit), Intel Core i5-7500 CPU @ 3.40 GHz, 16.0 GB RAM.
- Computer 2: Sequoia 15.6, Apple M4 Pro chip (12-core CPU, 16-core GPU, 16-core Neural Engine), 24.0 GB RAM.

➢ *Implementation Process*

The system implementation followed a four-stage workflow:

- Collection of image datasets for target objects.
- Training the AI model using the YOLO framework for object classification.
- Integration of scripts into the Visual Studio Code environment. 4) Program execution and experimental evaluation.

➢ *Detailed Implementation Steps*

- *Step 1: Image Acquisition and Analysis*

Initial data collection focuses on capturing environmental images to facilitate the separation of human subjects from other objects.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

Fig 16 Images Stored for Object Analysis

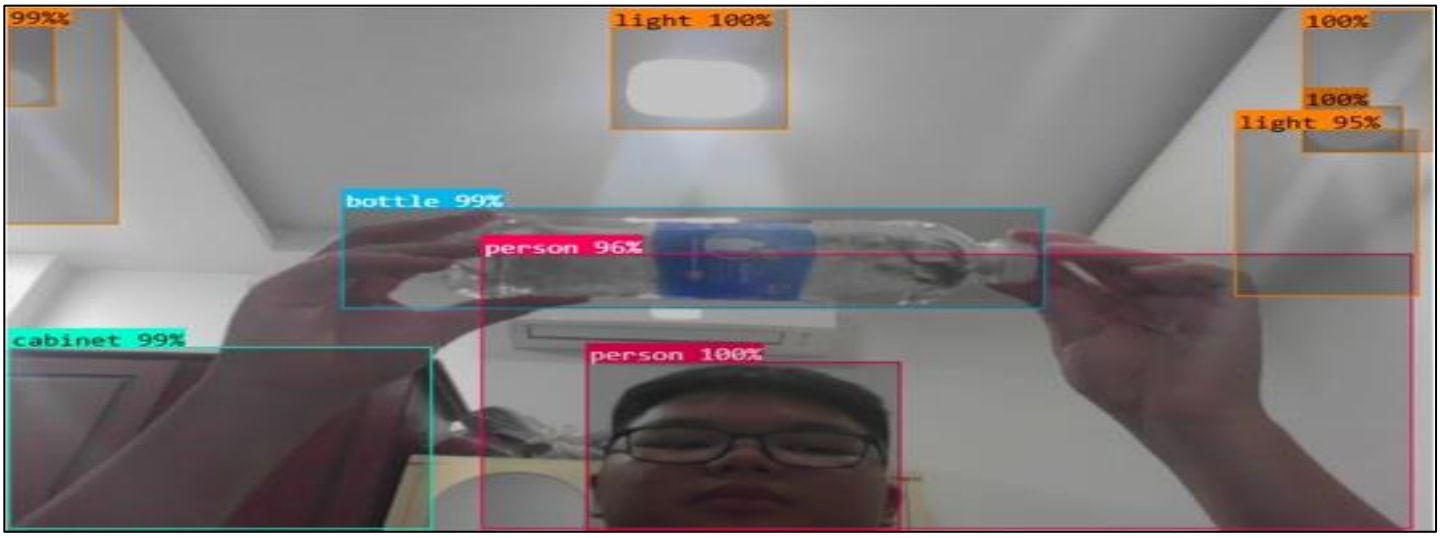

Fig 17 Object Recognition Results Illustrated by AI


Fig 18 Object Recognition Results Illustrated by AI

- *Step 2: Model Training*

AI training is performed in Visual Studio Code using pre-developed algorithms. The system analyzes object features to build a robust recognition model.

- *Step 3: System Output*

Final validation is achieved by running the program and observing the real-time classification results generated by the inference engine.

➢ *Experimental Results*

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

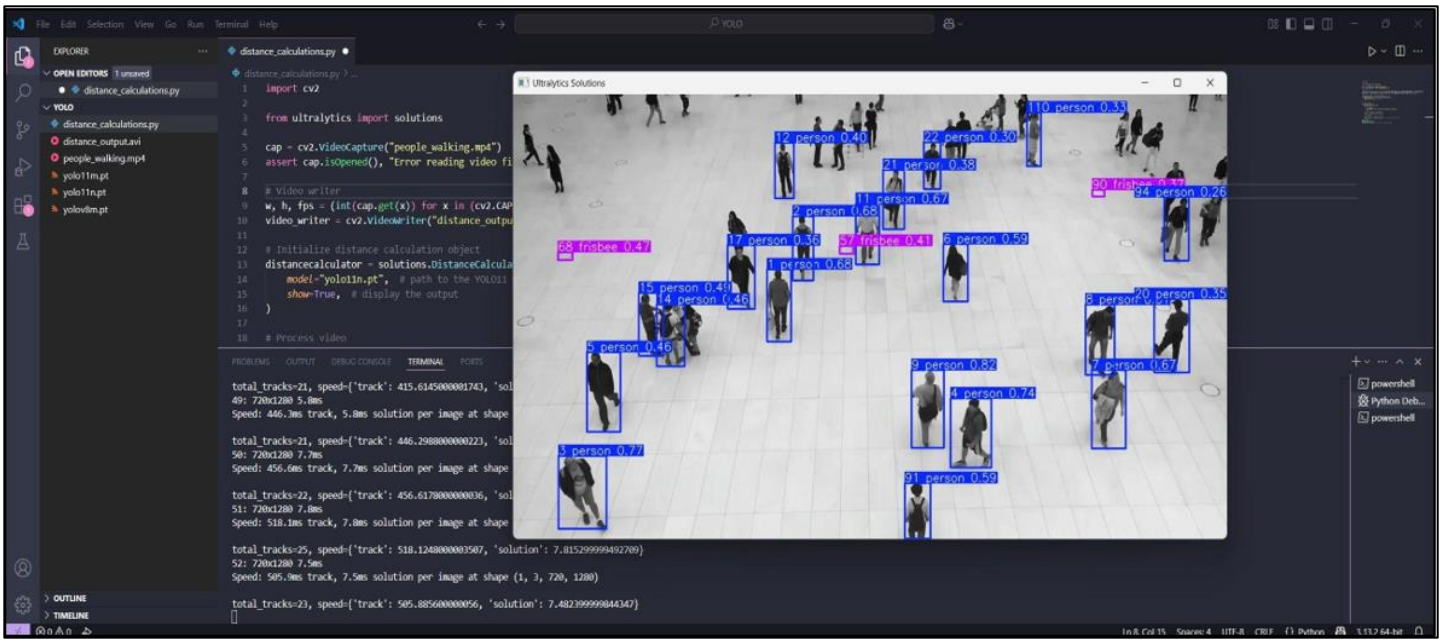

Fig 19 Human Detection Results After AI Training

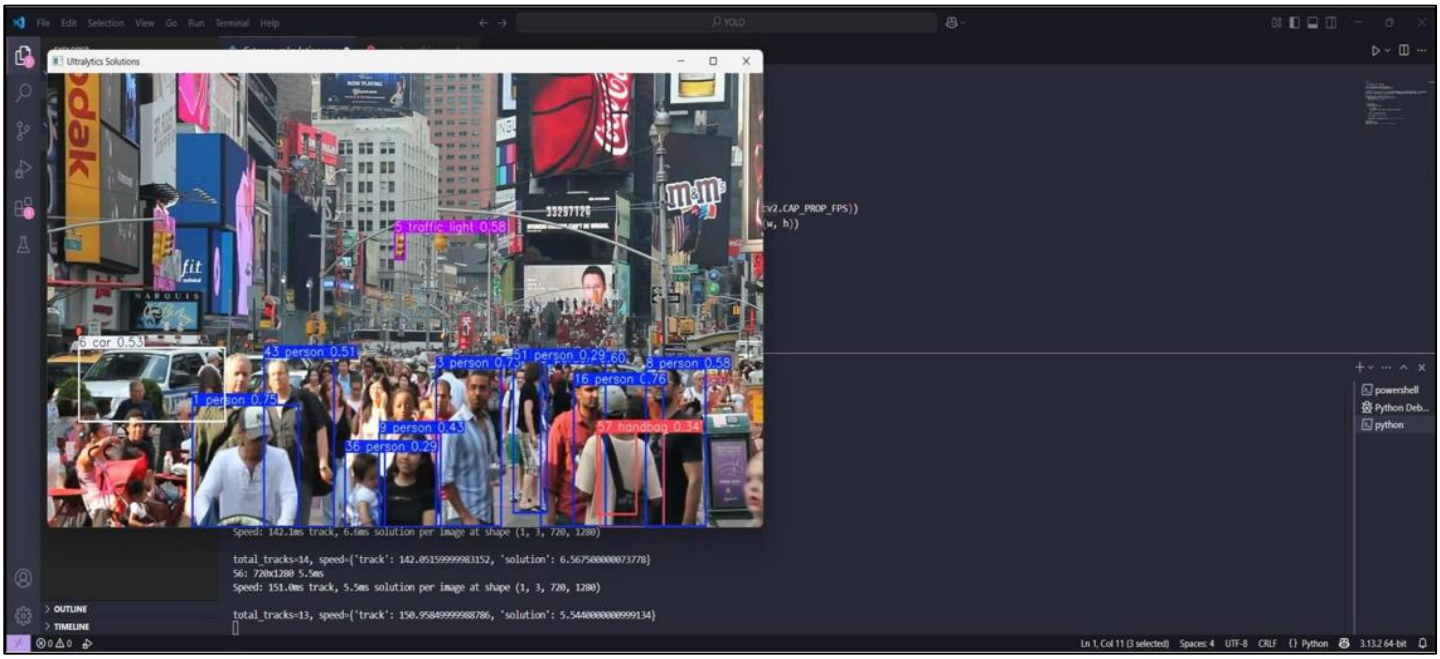

Fig 20 Human Detection Results After AI Training

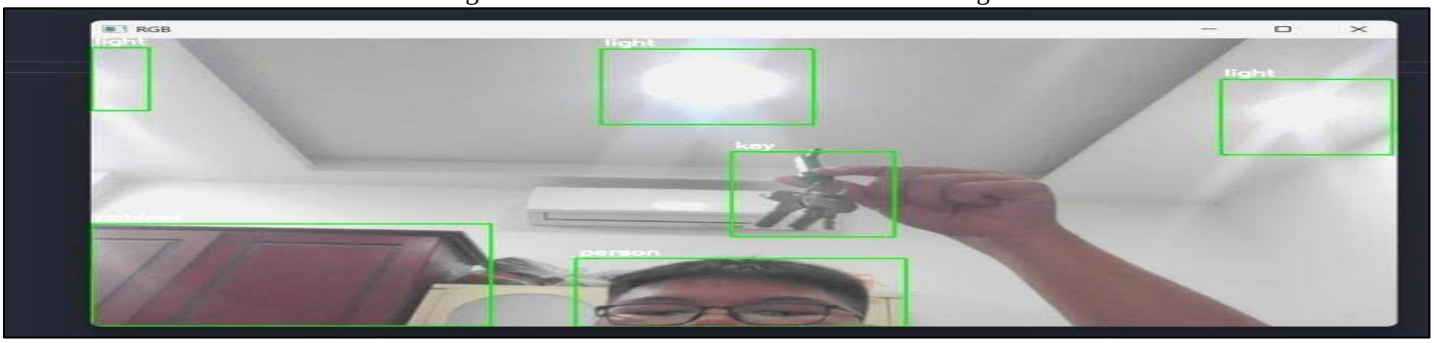

Fig 21 Detection of Trained Objects

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

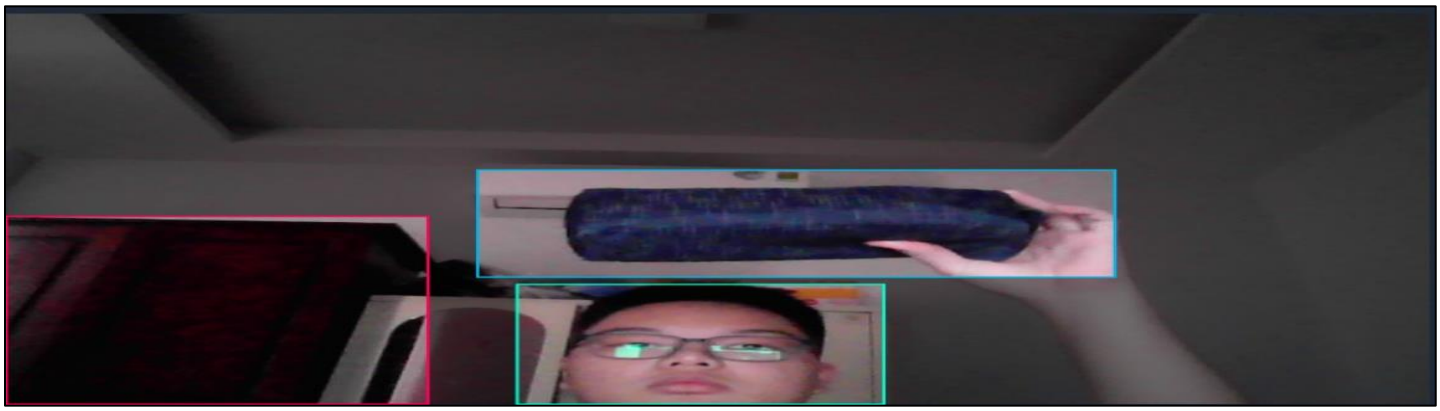

Fig 22 Distance Measurement Between the Object and the Camera

Fig 23 Illustration of Distance Measurement from the Object to the Camera

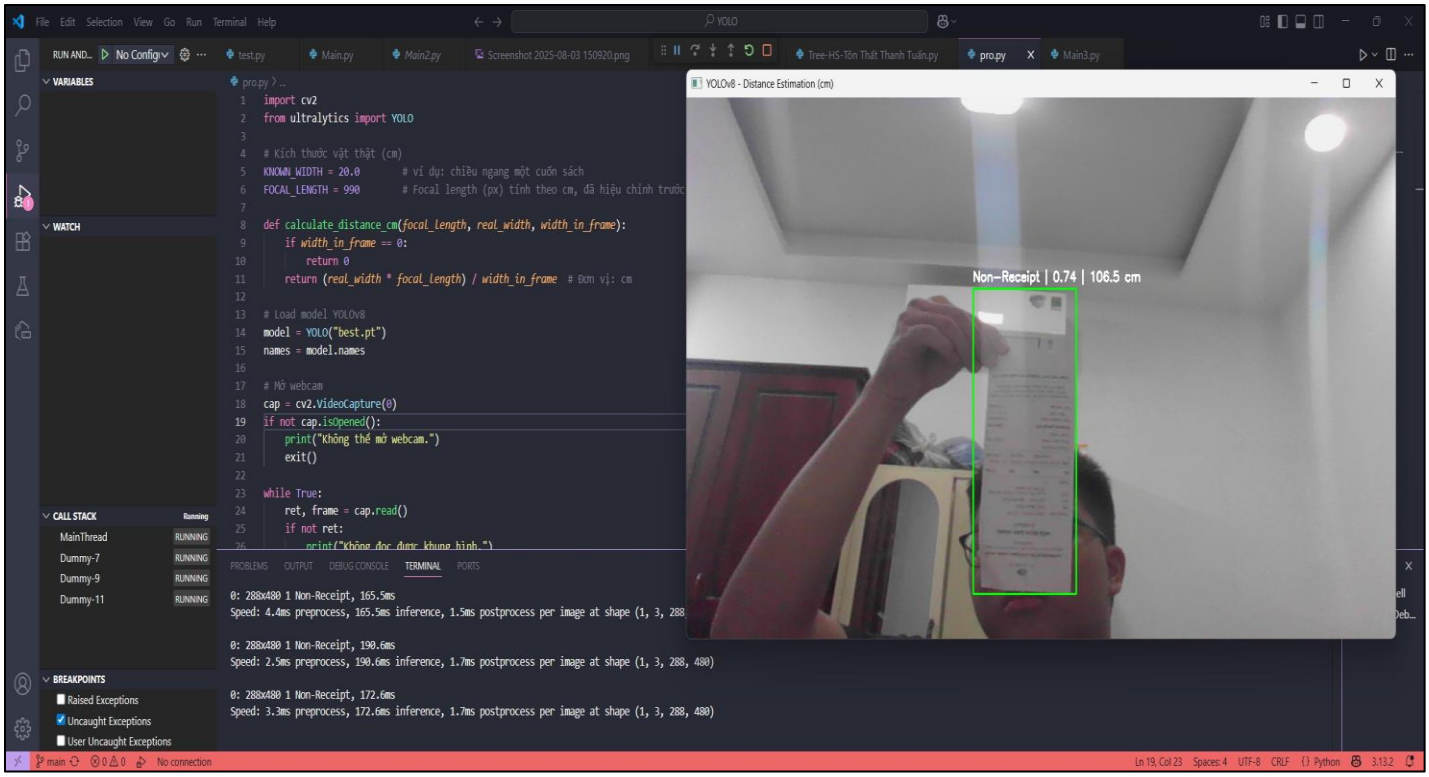

Fig 24 Illustration of Distance Measurement from the Object to the Camera

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

➢ *Evaluation in Stable Environmental Conditions*

The evaluation process was conducted under two conditions: a stable environment and a lowlight environment, in order to obtain a more objective assessment of the system.

In this experiment, the dataset was divided using the hold-out method, in which the data is split into two independent sets: a training set and a testing set. To avoid class imbalance in the training dataset, samples were distributed evenly between the two sets.

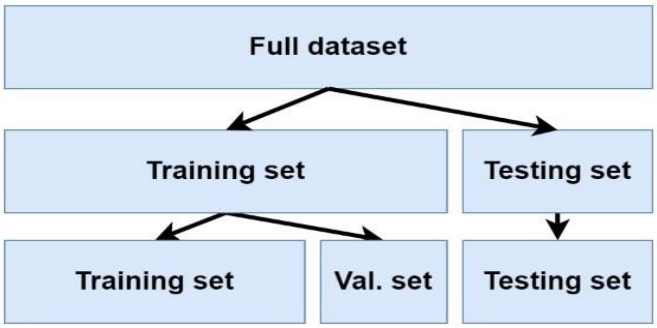

Fig 25 Dataset Division Using the Hold-Out Method [4]

➢ *Input Data in Stable Environmental Conditions*

For the object and human recognition system, the team utilized a dataset constructed based on the You Only Look Once (YOLO) framework. To facilitate model development, 70% of the data was allocated for training, while 30% was reserved for validation and testing under stable environmental conditions. The specific distribution of images across various classes is detailed in Table 7.

Table 7 Dataset Distribution for YOLO-based Recognition System

Class	Total Images	Training	Validation	Testing
Person	20	14	4	2
Bottle	7	4	2	1
Cabinet	10	7	2	1
Key	7	4	2	1
Light	80	56	12	12
Pencil case	5	3	1	1
Remote	4	2	1	1
Total	133	90	24	19

➢ *Experimental Results of Object and Human Detection in a Stable Environment*

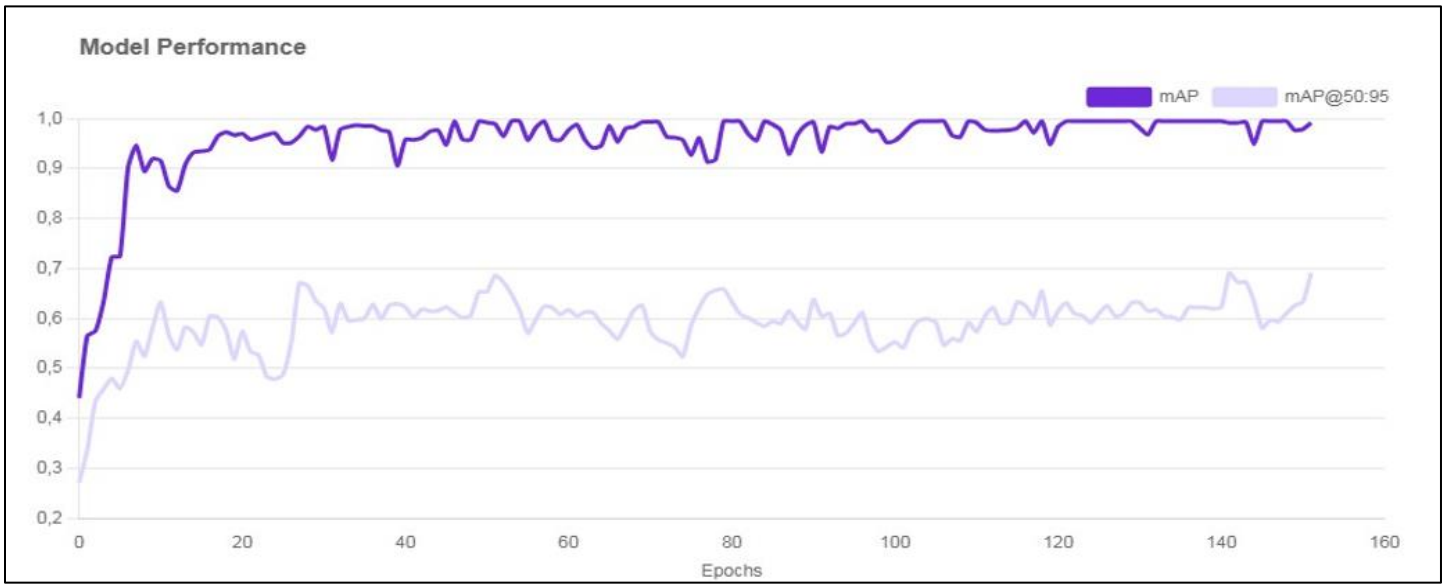

Fig 26 Model Performance Chart in a Stable Environment

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

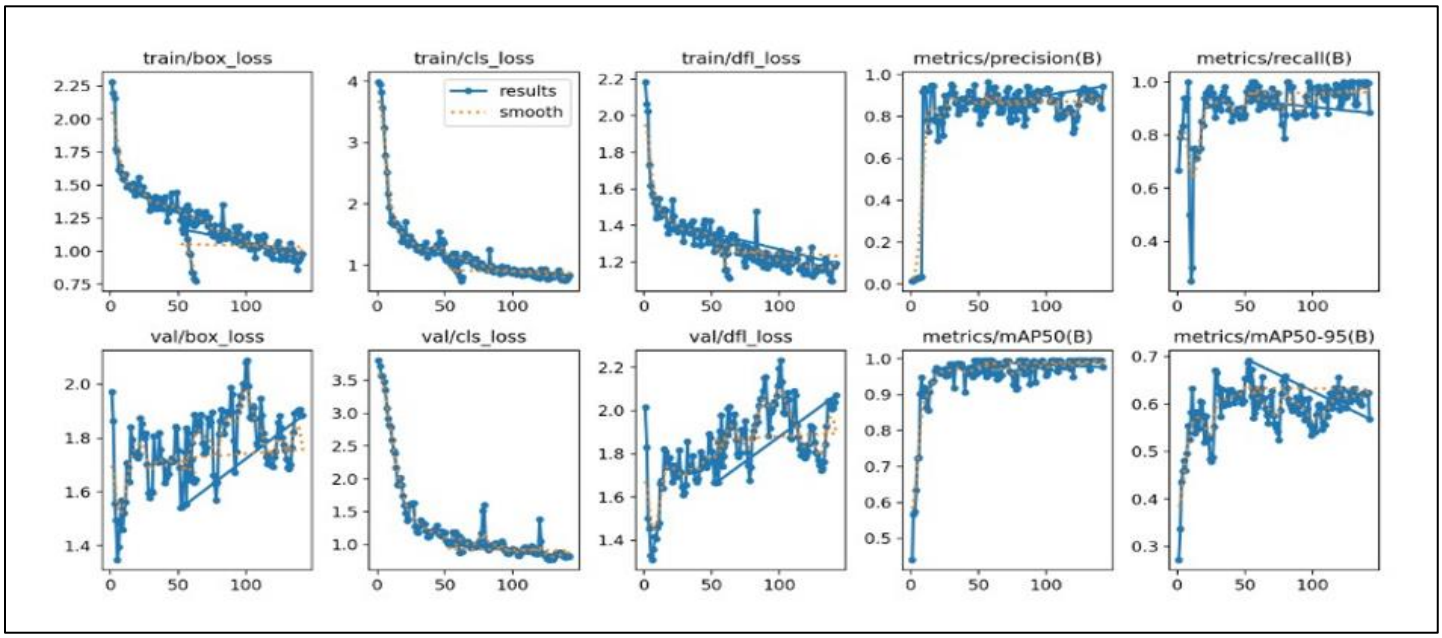

Fig 27 Improved Performance Chart After AI Training in a Stable Environment

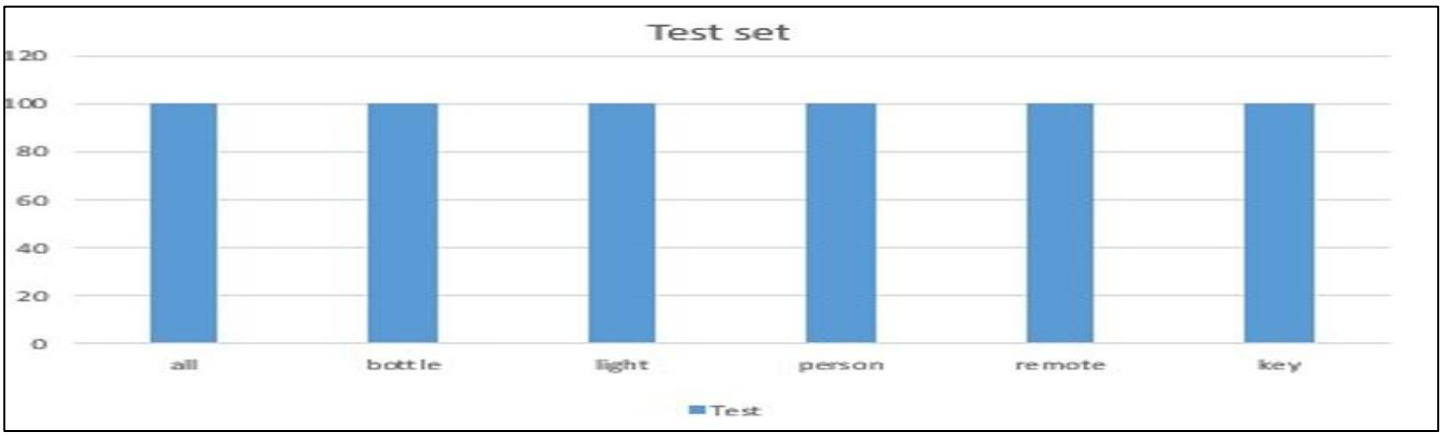

Fig 28 Statistics of Object Detection Accuracy After Testing in a Stable Environment

Fig 29 Statistics of Object Detection Accuracy After Validation in a Stable Environment

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

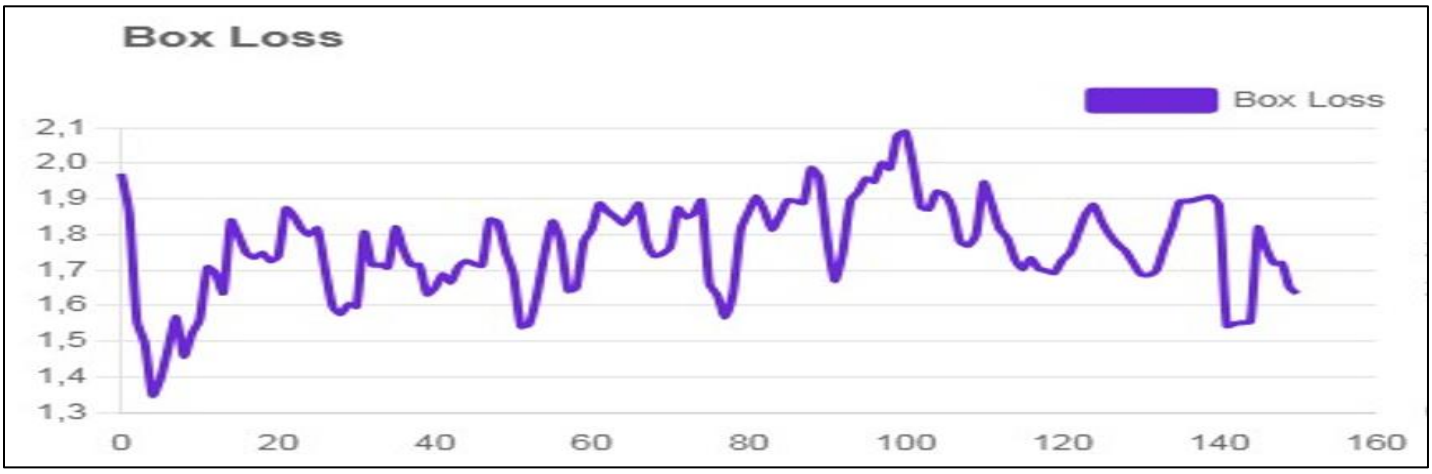

Fig 30 Chart Showing the Deviation Between Predicted and Actual Objects in a Stable Environment

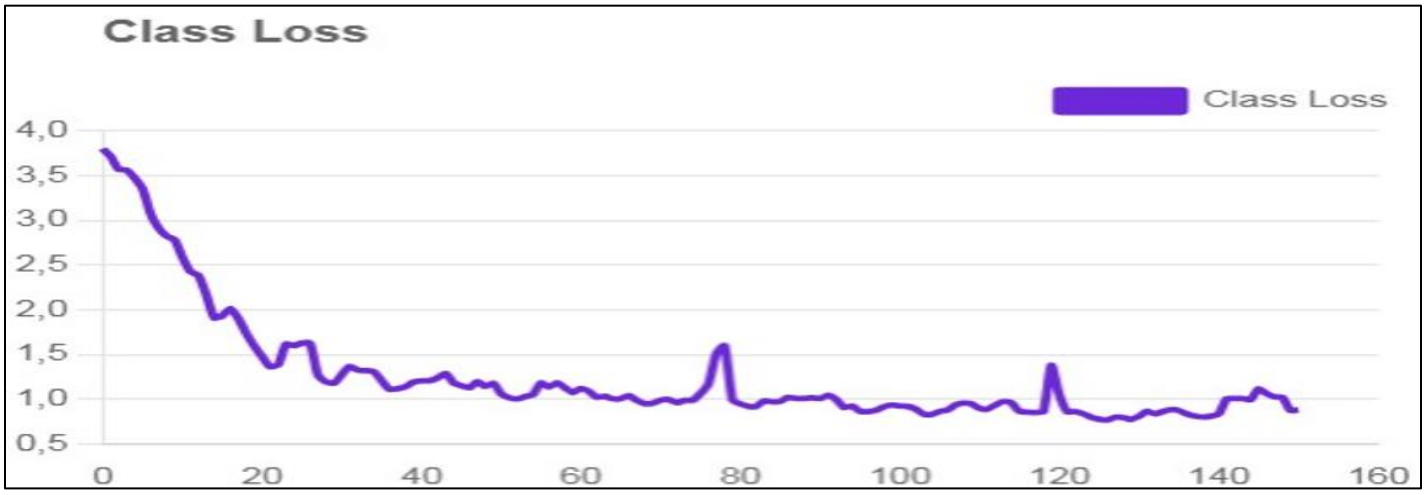

Fig 31 Chart Showing the Deviation Between Predicted Classification Labels and Actual Labels in a Stable Environment

➢ *Evaluation Process in Low-Light Environment*

Low-Light Feature Extraction and Preprocessing to mitigate the inherent challenges of lowlight environments—such as high sensor noise and reduced contrast—the system may incorporate an enhancement layer prior to detection. This ensures that the YOLO framework can extract meaningful features even when the pixel intensity values are concentrated in the lower end of the histogram.

Dataset Composition and Class Distribution The dataset encompasses a diverse range of nocturnal and indoor low-illumination scenarios to prevent model overfitting to specific lighting artifacts. The class distribution reflects common obstacles and subjects encountered in the system's target environment.

Training Configuration The training process utilizes a stochastic gradient descent approach with a momentum-based optimizer. To adapt the YOLO framework to the 30% validation/test split, we apply heavy data augmentation—including random flipping and synthetic noise injection—to simulate the thermal noise typical of high-ISO photography. This ensures that the 70% training subset remains robust when deployed in unpredictable real-world nighttime conditions. 8.

Table 8 Dataset Distribution for Low-Light Performance Evaluation

Class	Total Images	Training	Validation	Testing
Person	50	20	15	15
Bottle	17	7	5	5
Cabinet	48	20	15	13
Pencil case	14	6	4	4
Remote	19	8	6	5
Total	148	61	45	42

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

➢ *Experimental Results of Object and Human Detection in a Unstable Environment*

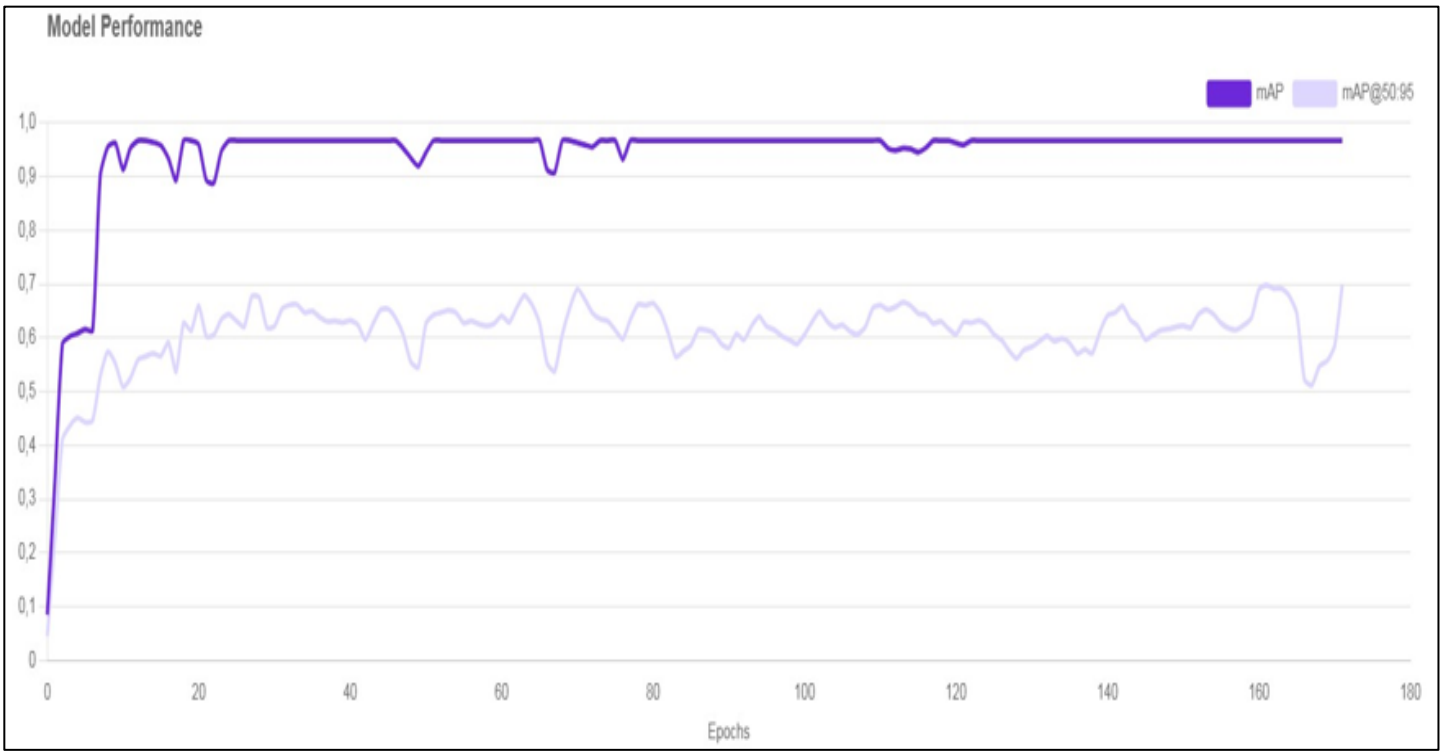

Fig 32 Model Performance Chart in a Unstable Environment

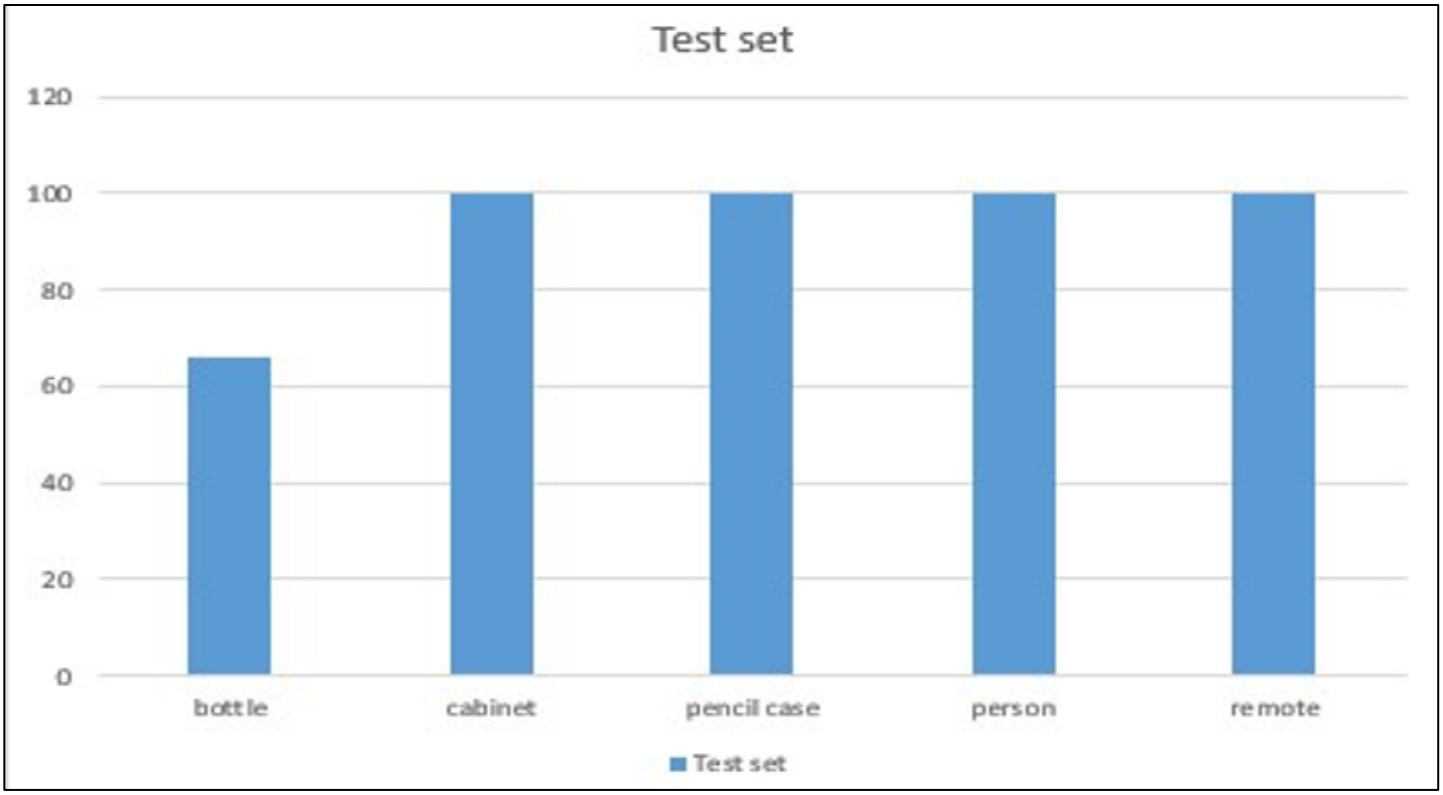

Fig 33 Improved Performance Chart After AI Training in a Unstable

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

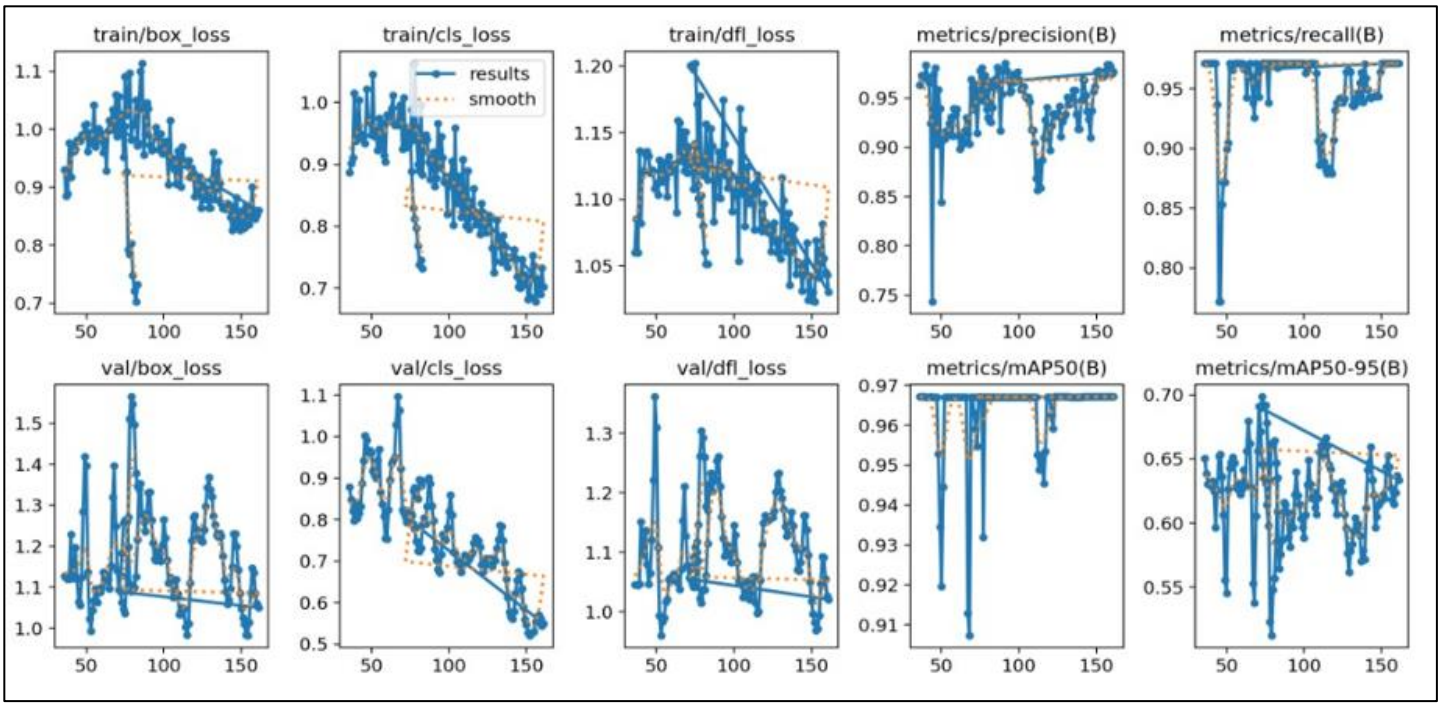

Fig 34 Statistics of Object Detection Accuracy After Testing in a Unstable Environment

Fig 35 Statistics of Object Detection Accuracy After Validation in a Unstable Environment

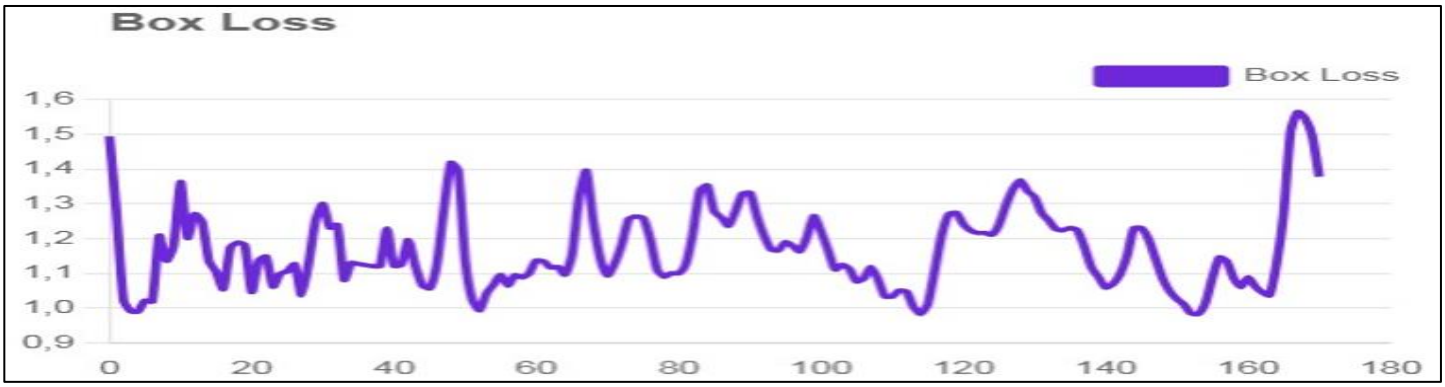

Fig 36 Chart Showing the Deviation Between Predicted and Actual Objects in a Unstable Environment

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

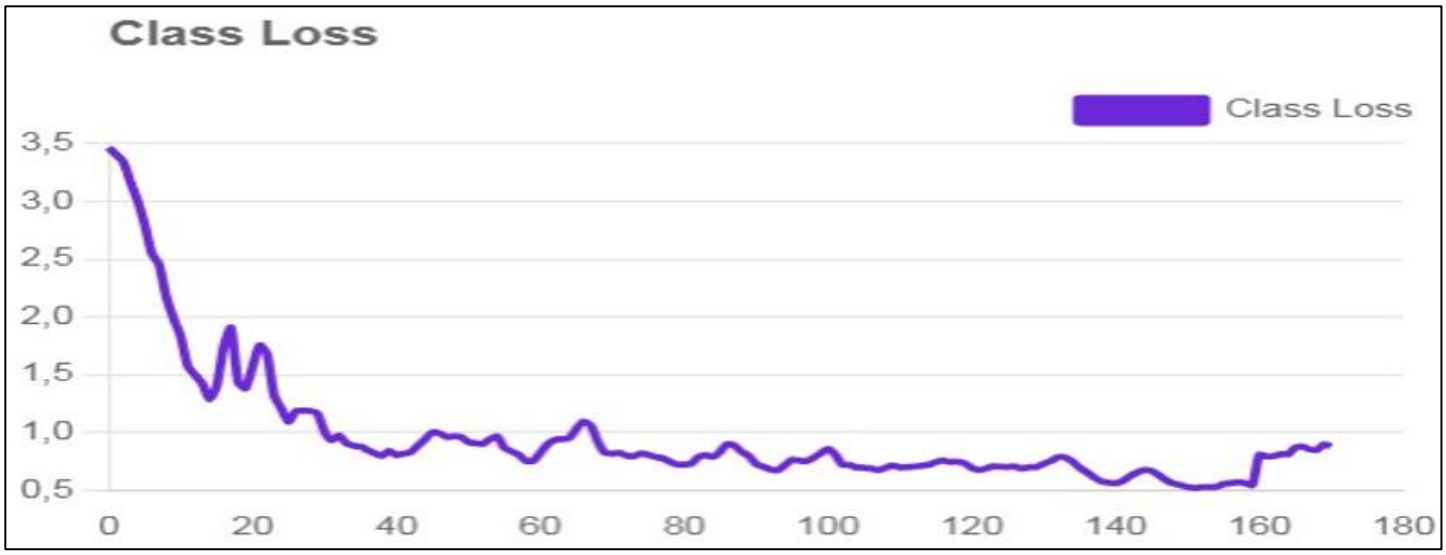

Fig 37 Chart Showing the Deviation Between Predicted Classification Labels and Actual Labels in a Unstable Environment

XIII. EXPERIMENTAL RESULTS COMPARISON

After conducting experiments in both stable and low-light environments, the following observations were obtained from the performance charts:

- *Model Performance:*
  In a stable environment, the model performance fluctuated between 90% and 100%. Meanwhile, in low-light conditions, the performance remained consistently around 100%.

- *Object Detection (Testing Set):*
  In a stable environment, detection accuracy reached 100% across all objects. However, in low-light conditions, performance dropped significantly for certain objects, notably the bottle class, which achieved only 66%.

- *Object Detection (Validation Set):*
  In stable conditions, recognition accuracy ranged from 98% to 100%. In contrast, the low-light environment showed inconsistency, particularly with the bottle class achieving only 85%.

- *Prediction vs. Ground Truth Error:*
  In stable environments, the error ranged from 1.6 to 2.1. In low-light conditions, the error remained lower, between 1.0 and 1.6.

- *Classification Label Error:*
  In stable environments, the error decreased from 4 to 1. In low-light environments, performance fluctuated between 3.5 and 0.

- *Object Presence Prediction Accuracy:*
  In stable environments, prediction density ranged from 1.7 to 1.9, while low-light environments showed lower values between 1.0 and 1.2.

Table 9 Comparative Analysis of System Performance

Metric	Stable Environment	Low-Light Environment
Model Performance	90% – 100%	≈ 100%
Testing Accuracy (Bottle)	100%	66%
Validation Accuracy (Bottle)	98% – 100%	85%
Prediction Error	1.6 – 2.1	1.0 – 1.6
Classification Error	4.0 → 1.0	3.5 → 0.0
Presence Prediction	1.7 – 1.9	1.0 – 1.2

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

XIV. CHAPTER 4. CONCLUSION AND FUTURE DEVELOPMENT

➢ *Conclusion*

This study presents the application of a modern object detection model, YOLOv11, in the development of an AI-based assistive system designed to support visually impaired individuals and elderly users in daily mobility. The proposed system enables real-time environmental perception by detecting surrounding objects and potential obstacles, thereby contributing to safer and more independent navigation.

Experimental observations indicate that system performance is strongly influenced by lighting conditions, with stable illumination significantly improving detection accuracy compared to low-light environments. This finding highlights the importance of environmental factors in the practical deployment of vision-based assistive technologies.

Beyond its technical implementation, the proposed approach demonstrates the potential of artificial intelligence to address real-world accessibility challenges. By translating object detection capabilities into functional mobility support, the system offers a foundation for assistive solutions that enhance autonomy and safety for vulnerable populations.

Future work will focus on improving robustness under varying environmental conditions and integrating additional intelligent features to further expand system functionality and real-world applicability.

➢ *Future Development*

The system has the potential for significant impact across several domains:

- *Navigation and Mobility:*

The system can help visually impaired individuals and older adults engage with modern life more safely and conveniently, especially in navigation and mobility within their surroundings.

- *Risk Mitigation:*

It can reduce the risk of accidents caused by collisions with obstacles, thereby improving quality of life and user independence.

- *Scalability:*

The solution has strong potential for large-scale deployment, particularly in urban areas, healthcare centers, or integration into smart wearable devices such as earmounted assistive technology.

- *Innovation:*

It contributes to promoting the application of science and technology in real-life contexts, opening new pathways for assistive healthcare technologies and human-centered innovation.

REFERENCES

[1]. Jan Egger, Christina Gsaxner, Xiaojun Chen, Jiang Bian, Jens Kleesiek, and Behrus Puladi. Apple vision pro for healthcare:" the ultimate display. *arXiv preprint arXiv: 2308.04313*, (2), 2023.

[2]. Seth R Flaxman, Rupert RA Bourne, Serge Resnikoff, Peter Ackland, Tasanee Braithwaite, Maria V Cicinelli, Aditi Das, Jost B Jonas, Jill Keeffe, John H Kempen, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, 5(12):e1221–e1234, 2017.

[3]. JASD Fonseca, Antonio Baptista, Ma Joao Martins, and Joao Paulo N Torres. Distance measurement systems using lasers and their applications. *Applied Physics Research*, 9(4):33–43, 2017.

[4]. Tung Sum Thomas Kwok, Zeyong Zhang, Chi-Hua Wang, and Guang Cheng. Towards high supervised learning utility training data generation: Data pruning and column reordering. *arXiv preprint arXiv:2507.10088*, 2025.

[5]. Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.

[6]. Michael Moll. Displacement damage in silicon detectors for high energy physics. *IEEE Transactions on Nuclear Science*, 65(8):1561–1582, 2018.

[7]. Nitin Rane. Yolo and faster r-cnn object detection for smart industry 4.0 and industry 5.0: applications, challenges, and opportunities. *Available at SSRN 4624206*, 2023.

[8]. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[9]. Hector Rodr´ ´ıguez-Rangel, Luis Alberto Morales-Rosales, Rafael Imperial-Rojo, Mario Alberto Roman-Garay, Gloria Ekaterine Peralta-Penu˜ nuri, and Mariana Lobato-B˜ aez. Analysis´ of statistical and artificial intelligence algorithms for real-time speed estimation based on vehicle detection with yolo. *Applied Sciences*, 12(6):2907, 2022.

[10]. Manoj Sahni, Ritu Sahni, and Jose M Merig´ o.´ *Neural Networks, Machine Learning, and Image Processing: Mathematical Modeling and Applications*. CRC Press, 2022.

[11]. Mohammad Javad Shafiee, Brendan Chywl, Francis Li, and Alexander Wong. Fast yolo: A fast you only look once system for real-time embedded object

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26feb637

detection in video. *arXiv preprint arXiv:1709.05943*, 2017.

[12]. Akhilesh Sharma, Vipan Kumar, and Louis Longchamps. Comparative performance of yolov8, yolov9, yolov10, yolov11 and faster r-cnn models for detection of multiple weed species. *Smart Agricultural Technology*, 9:100648, 2024.

[13]. Do Thuan. Evolution of yolo algorithm and yolov5: The state-of-the-art object detention algorithm. 2021.

[14]. Tao Xu, Wei Shen, Xiaoshan Lin, and Yi Min Xie. Mechanical properties of additively manufactured thermoplastic polyurethane (tpu) material affected by various processing parameters. *Polymers*, 12(12):3010, 2020.

[15]. Wei You, Changqing Shen, Dong Wang, Liang Chen, Xingxing Jiang, and Zhongkui Zhu. An intelligent deep feature learning method with improved activation functions for machine fault diagnosis. *IEEE access*, 8:1975–1985, 2019.

[16]. Xing Zhang, Gongjian Wen, and Wei Dai. A tensor decomposition-based anomaly detection algorithm for hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):5801–5820, 2016.