Object Detection Based on You Look Only Once Version 8 for Real-Time Applications

Gede Agus Santiago[1], Putu Sugiartawan*[2], Ni Nengah Dita Ardriani[3]
[1]Program Studi Rekayasa Sistem Komputer, Institut Bisnis dan Teknologi Indonesia, Bali
[2,3]Program Studi Teknik Informatika, Institut Bisnis dan Teknologi Indonesia, Bali
e-mail: [1]agus.santiago@instiki.ac.id, *[2]**putu.sugiartawan@instiki.ac.id**,
[3]dita.ardiari@instiki.ac.id

Abstract

This research focus to involves human detection in crowded situations, especially in the lecturer's room. The lecturer's room is very vulnerable because it can be accessed by anyone with only one entry and exit to the lecturer's room, so it would be perfect to place this Yolo camera in front of the lecturer's room so that incoming and outgoing activities can be monitored during work days on campus. The main challenge is how the system can distinguish individuals in dense crowds and identify their relative locations to each other. In this context, it is necessary to find a solution that can overcome the uncertainty of recognizing individuals in a group and accurately understand the location and distance between them. One proposed solution is to use the YOLO algorithm on video recordings to detect human objects in the lecturer's room during working hours. This research introduces the YOLOv8 model, a real-time detection system with high speed and accuracy in detecting and classifying objects in video recordings. YOLOv8 can accurately detect object movement, making it an efficient real-time framework for dealing with complex objects. This research experiment involved using eight different smartphone devices to collect datasets. Using various smartphone devices aims to test object detection performance under various shooting conditions, including variations in image quality, lighting, shooting angle, and camera resolution. The research results show that using multiple smartphone devices in dataset collection can improve the robustness and accuracy of object detection models. By integrating datasets from various sources and shooting conditions, the YOLOv8 model was successfully trained to better recognize objects in different situations, even in campus environments that often have challenges such as weather variations and lighting fluctuations. The test results show an accuracy rate of 93.33% in human object detection

Keywords— *Object detection, human detection, YOLOV8, Accuration*

1. INTRODUCTION

This Humans see images, immediately process the objects in them, and determine their location due to interconnected brain neurons. The human brain is very accurate in performing complex tasks, such as identifying objects with similar attributes, quickly. Like human interpretation, today's world requires fast and accurate algorithms to classify and detect objects for various applications. In recent years, methods based on deep learning have experienced rapid development in achieving high levels of accuracy. One popular strategy is You Only Look Once (YOLO). YOLO is a real-time object detection approach capable of detecting and classifying objects in videos quickly and efficiently[1]. Classifiers are used for image classification and counting. The object classification process and detection workflow aims to classify objects based on their features and attributes. As time passed, many approaches were combined to get better results. Object detection approaches have evolved from sliding window-based methods to single-shot detection frameworks. Detection and classification of objects in video involves processing each frame sequentially in YOLO v8. YOLO v8 can recognize and classify various object categories, such as humans, cars, bicycles, animals, etc. Once an object is detected, YOLO v8 provides a bounding box that surrounds the object and provides an appropriate class label[2].

In activities in the campus environment, there is the potential to face significant problems, especially regarding access to human detection in crowded situations, especially in the lecturer's room. This smartphone camera is in front of the lecturer's room so that incoming and outgoing activities can be monitored during working days on campus. This situation can become even more complicated when suspicious activity or attempts at unauthorized access to campus areas occur. This problem can disrupt overall campus order and security, so it is necessary to take practical preventive and supervisory steps to maintain the campus environment's integrity and protect all academic community members from potential risks and threats. Using smartphone cameras using the YOLO (You Only Look Once) method in campus environments significantly improves security and surveillance. By implementing YOLO on the camera, the system can accurately detect the presence of human objects in real-time. This will help identify and record suspicious activity or unauthorized access to campus premises.

The You Only Look Once (YOLO) method detects objects quickly and accurately. This method can detect objects up to 2 times faster than other algorithms. YOLO v8 combines the basic ideas of YOLO with various improvements in accuracy and performance. YOLO v8 uses a deep convolutional neural network architecture, which is often called Darknet [3] Darknet architecture uses multiple convolutional layers to obtain rich feature representation from images/videos. The YOLO model follows a specific flow method to analyze and detect objects quickly It follows a regression model where it takes input and derives class probabilities calculate class-specific confidence scores It compares the confidence score with a predetermined threshold value for detecting and classifying objects. If the confidence score exceeds the threshold value, the algorithm will not detect that particular object [4] This study, the YOLOv8 model is introduced, a real-time detection system for detecting and classifying objects in video recordings. YOLO v8 is one of the latest object detection methods and has been proven effective in detecting and classifying objects in videos with high speed and good accuracy. This method uses a simultaneous regression approach that allows object detection in real time and can recognize different categories of objects in videos.

2. METHODS

This research involved using ten cell phones to run scenarios based on time intervals starting from working hours. YOLO v8, an Object Detection algorithm that allows human identification in each video frame, is used in calculating the number of people. The analysis begins by recording traffic video, then using the YOLO v8 algorithm to recognize human objects and provide labels and bounding boxes to show their position and identity. Next, the number of objects is calculated by calculating the total labels or bounding boxes for humans in the recording. The concept of zones or border lines is used to separate human objects entering and leaving a zone. This system monitors the central coordinates of moving human objects and counts humans crossing the entry and exit zones. Integrating the YOLO v8 method and movement analysis allows the system to provide detailed information about the total number of objects and the direction of movement in video recordings. The research will be carried out through several stages, namely collecting recorded video data using eight different types of cellphones, reading and marking frames, changing frame sizes, object detection algorithm (YOLO), thresholding, providing labels, counting the number of objects and producing detection results as in Figure 1.1. The following is an explanation of the process analysis stages carried out in Detecting the Number of Two Human Objects Based on You Look Only Once (Yolo)

1. Video recording is taken from the Instiki Campus smartphone camera recording.
2. Reading and scanning frames is the process of reading frames in video recordings that are stored in the folder that has been created.
3. Changing the frame size is changing the frames from the default video frames file (1280 x 720) to 320 x 320.
4. Object Detection Algorithm (YOLO) is an object detection process using the Yolo algorithm version 8.
5. Thresholding is an image segmentation method that separates objects from the background based on differences in brightness or darkness.
6. Providing labels means giving labels or bounding boxes to objects in the video
7. Counting the number of objects is the process of calculating objects that have been labeled or bounding boxes.

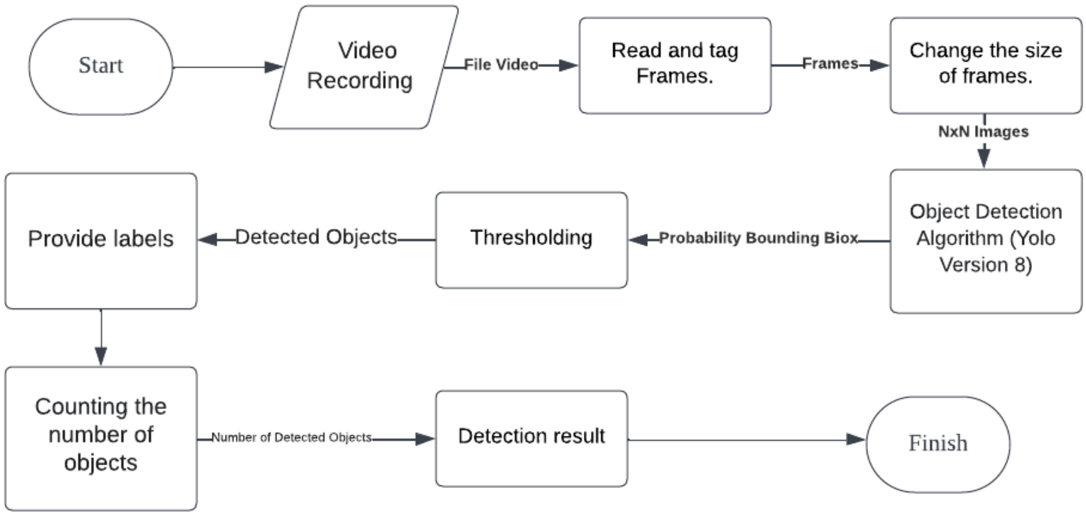

Figure 1. Research Stages

The flowchart above illustrates the steps in detecting objects in human video recordings, from capturing frames to labeling detected objects using the YOLO algorithm. Object detection results, including the number and type, can be used to optimize an object, increase security, and analyze human movement patterns. This information is also invaluable in the development of autonomous and intelligent technologies. With careful use, this object detection system has great potential to increase security efficiency in the campus environment and provide valuable insights for future development.

2.1 Object Detection

Object Detection or Object detection in digital image processing is a process used to determine the presence of particular objects in a digital image. The detection process can be carried out using various methods, which generally read the features of all objects in the input image. Object detection is a technique from Computer Vision (a field of artificial intelligence that discusses how machines can see like humans—one technique for finding objects in images or videos. Object detection also includes various Computer Vision tasks such as instance segmentation, Image Captioning (image naming), and Object Tracking (object tracking). Object detection is widely used in various fields of work, from the health sector to the security sector. For the transportation sector, object detection is usually used to detect objects that interfere with driving comfort [5], as in Figure 2.

Figure 2. Human image Detection of human objects with a camera

The reference In an era of ever-developing technology, the role of humans remains crucial. Even though we are drowning in a sea of information and technological sophistication, the human ability to think creatively, empathize, and collaborate remains unmatched. Humans are the main drivers of technological innovation, developing solutions to global challenges, creating inspiring content, and applying ethics in the use of technology. Even though technology makes our lives easier in many ways, humans must always balance technological progress and human values, making technology a tool that supports human development.

Object detection involves identifying the presence of objects, object coverage, and determining location in images. This object detection process functions to recognize object classes in the trained database. One of the current methods in object detection, such as YOLOv8, enables automatic and accurate detection of human objects in various visual contexts. YOLOv8 has delivered several significant innovations, along with flexibility, making it a desirable choice for a wide range of object detection and image segmentation tasks. In the context of human object detection, YOLOv8 uses the Anchor Box technique to identify possible locations of objects appearing in images or video recordings. This model partitions the image into small grids and identifies human objects in each grid with the help of an object-based approach.

2.2 Realtime Video

The video comes from Latin, precisely from the word "video" or "visum," which refers to seeing or having the ability to see. Video is a collection of still images arranged in such a way that they create movement when displayed in a frame at a certain speed[6]. This occurs in a different dimension from the original object. Live shooting and presentation over the internet, also known as live broadcasting or live broadcasting, is a way to control, record, and process data so quickly that the results can be received almost simultaneously. The use of the Faster R-CNN method allows trash detection in real-time videos. The real-time video involves manipulating, recording, and processing data at high-speed levels, allowing for near-real-time viewing. In this process, moving images are arranged in frames with different dimensions and presented in real-time through technology such as live broadcasting. In other words, real-time video provides a visual experience simultaneously with the captured or broadcast event[7]

2.3 You Only Look Once Version 8 (YOLO V8)

You Only Look Once (YOLO) is a deep learning-based object detection algorithm first proposed by Redmon et al. YOLO is an object recognition algorithm derived from developing the CNN (Convolutional et al.) method. This algorithm is the first “One Step Detector” object detection algorithm based on deep learning. As the name suggests, the You Only Look Once (YOLO) algorithm only uses a neural network at one level on the image. This network divides the image into several regions while predicting the bounding box and probability of each region [8].

Figure 3. YOLO Algorithm Model

2.3.1. YOLO Architecture Version 8

The YOLO (You Only Look Once) architecture is a prevalent approach to image processing and object detection in computer vision. YOLO is designed to detect and recognize different objects in an image in real time with excellent performance. Compared with traditional object detection approaches, YOLO carries the concept of "single-stage" detection, which means that object detection and classification are carried out in one single stage. This approach differs from the "two-stage" approach, which usually involves a proposal detection stage (for example, using methods such as R-CNN and Faster R-CNN) followed by a classification stage. YOLO can predict objects directly on the entire image, simultaneously producing bounding boxes and class probabilities. With an architecture consisting of convolutional layers, YOLO enables faster information processing, making it a popular choice in real- time applications such as object detection in autonomous vehicles and security systems. Although there are various versions of YOLO with performance improvements, the core of the YOLO architecture remains focused on the ability to detect objects accurately and efficiently.

2.3.2 Evaluation of YOLO Version 8

YOLO (You Only Look Once) evaluation is a process to measure the performance and accuracy of the YOLO object detection model in carrying out detection tasks on specific datasets. Evaluation is carried out to assess how the model can recognize objects correctly in various situations and environments. The model measures precision, namely the ratio between objects that are predicted correctly compared to the overall results predicted by the model; recall, namely the ratio between objects that are predicted correctly compared to the overall actual results, mean average precision (mAP) is calculated using precision and recall values. Precision(P), recall(R), average precision(AP), and mean average precision(mAP).

3. RESULTS AND DISCUSSION

This section contains In this section, the focus shifts to the presentation and discussion of the results of the experiments that have been carried out. This section will comprehensively explore these findings, providing a detailed analysis of the data obtained during the testing phase. The aim is to provide insight into the performance and effectiveness of the YOLOv8 algorithm in detecting objects on various cellphone brands and camera specifications. This section will not only reveal the testing process results but will also engage in a thorough discussion to interpret the significance of those results. Additionally, it will address any observed variations in detection accuracy, speed response, and the algorithm's ability to identify objects in varying lighting conditions. This review will present a thorough understanding of the algorithm's performance in object detection, contributing to a deeper understanding of its advantages and limitations.

3.1 Data testing

The test data consists of a series of images taken with the various cellphone brands mentioned previously, such as the Samsung A71, Redmi 9A, Redmi Note 12, iPhone 11, Redmi Note 9, Realme C17, Poco M3, and Vivo Y21, Redmi 8, iPhone 11 pro max. Each cellphone has different camera specifications, and the images produced by the cellphone are analyzed using the YOLO algorithm to evaluate the level of object detection by this algorithm on each device. Evaluating the detection accuracy, response speed, and recognition of YOLO objects under different lighting conditions and backgrounds will be the main focus of this test. By involving various popular mobile phone brands, this test aims to provide an in-depth understanding of YOLO's performance on these devices, enabling a comprehensive assessment of the object detection algorithm's capabilities in different photographic contexts, show in Table 1.

3.2 Object Detection Results with YOLOv8

The performance of the YOLO v8 model will be evaluated using a confusion matrix, recall, precision, and F1 score on a previously trained human detection network. The performance of both trained models will be evaluated using ten frames from the test set selected from the dataset. These ten frames cover various conditions in the campus area, from minimal activity to hectic situations. This test is designed to test the model's capabilities in multiple contexts and levels of complexity by taking samples from the ten available campus area records. Network performance will be tested by comparing actual human totals with totals predicted by the model using the author's dataset, including ten video data frames, as outlined in Table 2.

Table 1 Testing Data for Total Vehicle Detection

Phone	**Camera Specifications**	**Rear Camera VideoResolution**
Samsung A71	Quad Camera 64 MP, 12 MP, 5MP, 5 MP	1080p@30fps
Redmi 9A	13 MP rear camera and 5 MP frontcamera	1080p@30fps
Redmi Note 12	Three rear cameras (50 MP, 2 MP, 6 MP)	1080p@30fps
iPhone 11	12 + 12 MP rear cameras and 12MP front camera	4K@24/30/60fps, 1080p@30/60/120/240fps
Redmi Note 9	Main camera 48 MP with f/1.8 lensaperture	1080p@30fps
Realme C17	Quad camera with 13 MP mainlens and 8 MP selfie camera	1080p@30fps
Poco M3	Main camera 64 MP with 2 MP macro and depth cameras	1080p@30fps
Vivo Y21	Main camera 13 MP	1080p@30fps
Redmi 8	Main camera 12 MP	1080p@30fps
iPhone 11 Pro Max	Triple camera 12 MP + 12 MP + 12MP with 12 MP front camera	4K@24/30/60fps, 1080p@30/60/120/240fps

Table 2 Comparison of actual human totals and total human predicted results.

Device	**Video Resolution**	**Duration**	**Direct Observation**	**YOLOv8 Detection**
Samsung A71	1920 pixels to 1080 pixels	120 s	20	20
Redmi 9A	1080p at 30fps	120 s	17	17
Redmi Note 12	1080 pixels to 1920 pixels	120 s	15	15
iPhone 11	1080 pixels to 1920 pixels	120 s	21	21

Redmi Note 9	1080 pixels to 1920 pixels	120 s	17	17
Realme C17	1080p at 30fps	120 s	18	18
Poco M3	1080p at 30fps	120 s	15	15
Realme C17	1080p at 30fps	120 s	18	18
Poco M3	1080p at 30fps	120 s	15	15
iPhone 11 Pro Max	4K at 24/30/60fps, 1080p at 30/60/120/240fps	120 s	13	12

3.3 Data Testing with YOLO V8

Testing the data with YOLOv8 is an essential step in evaluating the performance of this object detection algorithm in a specific context. In this testing phase, we use datasets that cover various visual conditions, including differences in camera resolution, varying lighting conditions, and complex human detection object collection. Through this test, YOLOv8 will be assessed for its ability to recognize and provide bounding boxes for objects, especially human objects, accurately and consistently. We will also evaluate how the algorithm responds to differences in camera resolution, measuring its robustness and flexibility in dealing with varying visual conditions. The results of this test will provide in-depth insight into YOLOv8's ability to overcome challenges in object detection, contributing valuable insights to the INSTIKI campus environment effectively. The object detection results are presented in Figure 5 and Figure 6.

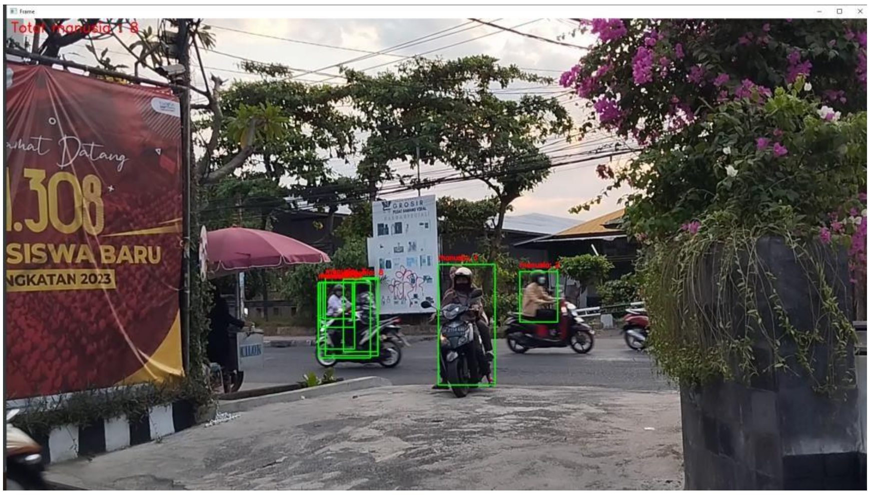

Figure 5. Object detection results using YOLO V8

Figure 5, depicting the results of human detection using a 50-megapixel resolution smartphone camera with the yolov8 algorithm, shows the advancement of technology in visual surveillance.

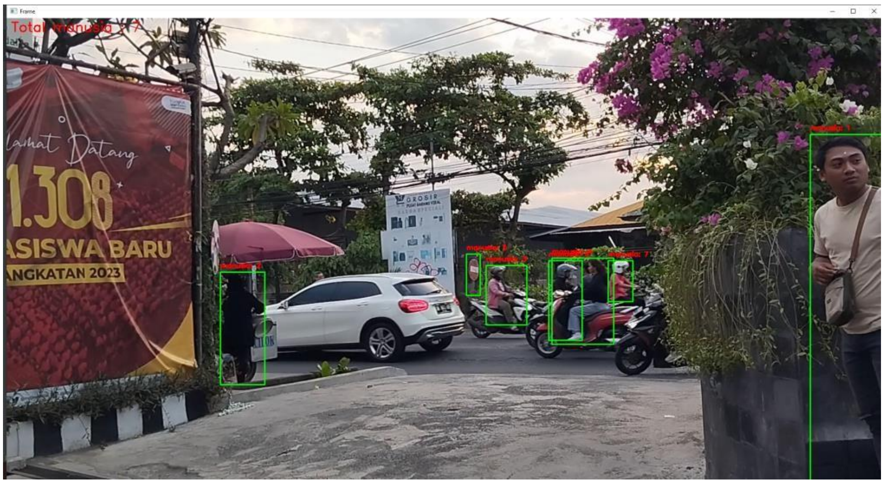

Figure 6.Object detection results using YOLO V8

although the image overall shows exceptional performance in object detection, there is one human object that was not detected by the algorithm. this lack of detection may be caused by several factors, such as human position, which may be obliged by other objects or insufficient lighting at the time the picture is taken. failure to detect one object not immediately reflects weaknesses in the overall algorithm. The image depicting the results of human object detection using a 1080-megapixel resolution smartphone camera with the YOLOv8 algorithm reflects the extraordinary level of precision and accuracy in visual monitoring. This figure shows that all existing humans are successfully detected by the algorithm, demonstrating YOLOv8's superior ability to recognize and provide bounding boxes for complex objects such as human objects. The 1080-megapixel high-resolution camera provides excellent visual detail, allowing the algorithm to identify each object and deliver consistent results effectively. The success of detecting all humans in this image illustrates the high level of accuracy of the YOLOv8 algorithm in handling variations in visual conditions and high-resolution camera settings

4. CONCLUSIONS

This research shows that applying human object recognition based on You Only Look Once Version 8 (YOLOv8) is an essential breakthrough in object recognition. With more accurate and faster detection capabilities, YOLOv8 offers an efficient solution for identifying human objects in various contexts. This technology allows the system to recognize human objects in real-time, providing advantages of security, monitoring, and other applications. With its innovative approach, YOLOv8 strengthens the foundation for developing future object detection technologies, enabling the software to adapt to complex environments.