Student Behavior Monitoring System in Classroom Environment Using YOLOv8

[1]Ifroh Intan Sepulau, [2]Kusumanto, [3]Nyayu Latifah Husni
[1,2,3]Program Studi D4 Teknik Elektro, Jurusan Teknik Elektro, Politeknik Negeri Sriwijaya, Indonesia
Email: [1]ifrohintan05@gmail.com, [2]manto6611@yahoo.co.id, [3]nyayu_latifah@polsri.ac.id

Article Info
Article history:
Received Apr 18th, 2025
Revised Jul 9th, 2025
Accepted Jul 30th, 2025

Keyword:
Deep Learning
Object Detection
Raspberry Pi
Sudent Ethics
YOLO

ABSTRACT
Student ethics in an academic environment is an important element in creating an orderly and professional learning environment. One form of ethical violation that is still often found in the lecture environment is eating and drinking activities in the classroom. This study's objective is to develop an automatic detection system for unethical student behavior in the classroom, especially eating and drinking activities, utilizing one of the newest Real-time deep learning approaches object recognition on a Raspberry Pi device, The algorithm known as You Only Look Once version 8 (YOLOv8). A special dataset was developed through a manual annotation process in the form of images and videos showing students with various activities in the classroom. This system is expected to be an additional solution in monitoring student ethics automatically, efficiently, and in real-time in a modern learning environment. The test findings demonstrate that the model can identify eating and drinking activities with a respectable degree of precision indicating that the system is able to detect target activities with an accuracy level of up 95% with fairly stable performance in good lighting conditions and certain viewing angles.

Corresponding Author:
Nyayu Latifah Husni,
Departement of Electrical Engineering,
Politeknik Negeri Sriwijaya,
Jl. Srijaya Negara, Bukit Lama, Bukit Besar, Kota Palembang, Sumatera Selatan, Indonesia, 30139
Email: nyayu_latifah@polsri.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v8i2.37086

1. INTRODUCTION

The classroom environment in the Electrical Engineering Department at Sriwijayan State Polytechnic is a place for students to learn and develop. In order for the teaching and learning process to run effectively, it is necessary to monitor the activities carried out by students as outlined in the rules and regulations when using and occupying the classroom. One form of ethical violation that often occurs but is considered trivial is eating and drinking in the classroom during lectures. This activity can disrupt the concentration of both instructors and other students, create an unfavorable learning environment, pollute the surroundings, and undermine the professionalism of the academic environment. With the increasing use of technology in education, an automated monitoring system based on Artificial Intelligence (AI) has become essential to help maintain order and ethical conduct among students in the classroom environment.

With the increasing use of technology in education, AI-based automatic monitoring systems have become important for helping to maintain order and ethics among students in the classroom environment, as well as for safeguarding and monitoring human activities in real time. Human-made computer vision-enabled systems are not very accurate or fast, even if a particular picture may be swiftly and precisely identified by the human eye, encompassing its position, content, and surrounding visuals that interact with it. Any developments in this area that boost productivity its performance could pave the way for the creation of systems with greater intelligence. that resemble people. Human life would then be made easier by these developments thanks to tools like assistive technologies, which enable people to do activities, for example with little or no conscious thought, even if the motorist is not conscious of what they are doing, operating a vehicle with computer vision-enabled assistive technology could anticipate and alert the driver to a collision before it occurs. Thus, real-time object detection has emerged as a crucial topic for the ongoing automation or substitution of human labor. Prominent areas of Machine Learning (ML) include computer vision and object identification, which are eventually anticipated to help unlock the potential of general-responsive robotic systems [1].

AI is one of the key areas of research in computer science. With its rapid technological advancement and vast area of application, AI is becoming pervasive very rapidly because of its strong applicability to issues, especially those that are difficult for people and conventional computing structures to handle [2]. So that it can automatically detect students' activities and behaviors in real time. Therefore, automated and real-time solutions such as the use of the YOLOv8 method are needed to quickly and accurately detect ethical violations, thereby maintaining the quality of education.

Natural object detection and language processing, voice/speech recognition, and image classification are the main subfields of deep learning. Specifically, there are two types of detectors for object detection: one-stage and two-stage, This uses an R-CNN-based model for A YOLO, SSD-based model and a two-stage detector for a one-stage detector. Among these, The model of the one-stage detector known as YOLO is incredibly quick and has a straightforward processing method, but its accuracy for small objects is rather low [3]. The You Only Look Once (YOLO) method was chosen because it has the main advantage of inference speed and computational efficiency compared to other deep learning methods such as Faster R-CNN and SSD. Compared to Faster R-CNN, which requires two processing stages (region proposal and classification), YOLO operates with a one-stage approach, making it faster and suitable for real-time applications. While SSD (Single Shot MultiBox Detector) is also a one-stage method, YOLOv8 has a lighter architecture, supports multi-scale detection with higher accuracy, and more precise bounding box processing thanks to the use of Distribution Focal Loss (DFL).

Additionally, YOLOv8 supports better detection of small objects through improvements to the backbone and neck architectures (such as the use of PANet and C2f modules). The ability to deploy on edge devices like Raspberry Pi is a significant advantage, as not all complex deep learning models can run stably on devices with limited computational power. As a result, YOLO is an ideal choice for implementation in student behavior monitoring systems that require fast, accurate, and efficient real-time detection.

Previous studies have widely explored object detection using deep learning methods, particularly the YOLO series. Redmon et al. (2016) [5] introduced YOLOv1, a one-stage object detector that unified classification and localization into a single network, enabling real-time detection performance. However, YOLOv1 struggled with small object detection and localization accuracy. To address this, Bochkovskiy et al. (2020) [6] developed YOLOv4 by integrating CSPDarknet53 as the backbone and advanced data augmentation strategies, resulting in improved accuracy while maintaining high processing speed. Lin et al. (2021) [7] applied the YOLO model in a smart surveillance system, demonstrating that YOLO is capable of detecting human activity in real time using camera feeds.

Previous studies have extensively applied YOLO-based methods for student behavior detection in classroom environments. Wang et al. (2023) proposed a YOLOv5-based model with a Squeeze-and-Excitation (SE) attention module and feature pyramid architecture, improving mAP by approximately 11% over YOLOv4 in recognizing multiple student learning behaviors (e.g., writing, listening) in real classroom setups [8]. Another study by Yang et al. (2023) introduced YOLOv7-BRA, which added Bi-level Routing Attention and model fusion to enhance detection performance of eight student behaviors (raising hand, reading, writing, etc.) in crowded scenes, achieving mAP@0.5 = 87.1% on the SCB-Dataset [9]. The present research extends these advancements by focusing specifically on detecting eating and drinking as ethical violations. Our system is deployed on Raspberry Pi for edge processing and tested in real classroom conditions. This combination of real-time detection, hardware deployment, and behavioral ethics context sets this study apart from previous works.

The suggested work serves as a tool for human monitoring, a security measure, and a channel for information sharing. The system aids in intrusion detection. Immediately after the Raspberry Pi PIR (Passive Infrared Sensor) senses movement, motion, or activity, the camera is activated. The latest model of Raspberry Pi4 has a camera that is shooting pictures. The invasion is located using image processing techniques such as the Simple Frame Difference Approach and Background Subtraction. An alert message containing a picture of the perpetrator will be sent to the mirror's administrator informing them of the intrusion [10]. The purpose of this study is to design a system for detecting student ethical violations in the classroom, especially eating and drinking, in order to create a clean and comfortable classroom environment using the deep learning method and to evaluate the effectiveness of YOLOv8 performance in the classroom environment in real time.

2. RESEARCH METHOD

Starting from the literature study, the author studies and collects sources manually from e-books, websites and looks for journal references on the internet and final project reports related to research. Looking for advantages and disadvantages of methods related to previous research. Planning the manufacture of hardware by determining the components needed to support the making of the project. This is done to find out the initial foundation to support the research report, then install and test the components to ensure everything works properly and according to specifications. Planning the development of software that will be used in the project, creating a design that includes program structure, algorithms, and user interface as well as writing program coding, implementing algorithms and developing features in accordance with the design made and testing to ensure the software works as desired. After that, integration and reporting of software and hardware are carried out to ensure that both can function properly by doing end-to-end to ensure the entire system runs optimally.

Dataset including pictures of pupils eating and drinking in class YOLO model training was conducted using after training, the model was integrated into a Raspberry Pi for testing in a real environment. Testing was conducted by simulating various scenarios, including students eating, drinking, and behaving normally in class.

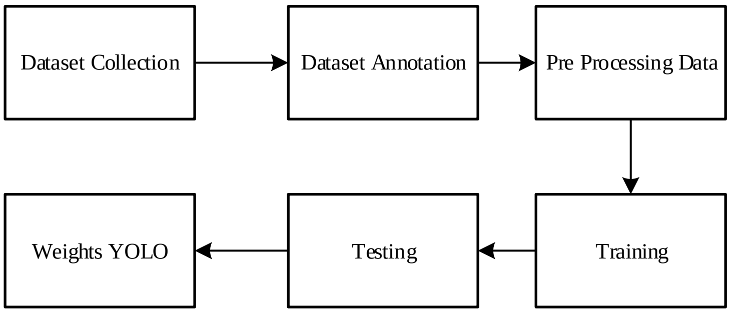


Figure 1. YOLO Diagram Blok

In Figure 1 the system process using YOLO begins with collecting datasets from several sources, one of which is roboflow, after collecting the next dataset, dataset labeling or annotation is carried out where this process is to determine what will be detected by YOLO, after labeling the next data pre-processing which includes the augmentation process including image rotation, Then the data training is carried out so that YOLO can detect the images we want, then the testing process to find out the weights to detect objects, if the model has gotten good testing we need to export data from YOLO to be used in the system that will be used.

2.1. Artificial Intelligence (AI)

An Intelligent Agent is a program that uses AI. An Intelligent Agent can engage with its surroundings. Through its sensors, the agent can determine the condition of its surroundings, and through its actuators, it can then modify that condition. The key component of AI is the agent's control strategy, which explains how sensor inputs are transformed into actuators that is, The mapping of sensors to actuators and is enabled by an internal function. The development of intelligence comparable to that of humans in robots is the ultimate aim of AI. Nevertheless, methods for learning that seek to imitate the way the the human brain is capable of learning achieve such a desire [11]. AI allows computers to carry out operations like speech recognition that formerly required human intellect, computer vision, and making decisions.

2.2. Machine Learning (ML)

Learning is the process of increasing one's knowledge or developing one's current skills, to put it simply. Learning key ideas and understanding their significance and how they relate to the topic at hand are just two of the procedures involved in acquiring new knowledge. From a biological perspective, skill improvement can be understood as strengthening a brain connection pattern that allows one to do the desired function [12].

Deep learning and ML are two rapidly developing subjects that have recently acquired attention. Both involve learning from data using algorithms to improve the accuracy and efficacy of predictions or decisions. While ML often uses statistical methods, Neural Networks are used in deep learning to extract knowledge from massive datasets. This article's the objective is to supply a broad overview of these technologies' methods and applications while also analyzing their benefits and drawbacks [13].

2.3. You Only Look Once (YOLO)

The rapid growth of deep learning algorithms has led to substantial breakthroughs in the object's field detection in recent years. There are two basic categories into which object detection falls approaches. Region-based two-stage detection models, which incorporate two separate procedures, are the first method. First, a list of possible object-containing candidate regions is suggested. The item categories inside each of these suggested regions are then identified by deploying a classification network on them. Mask R-CNN, which is based on masked areas, region-based Fully Convolutional Networks (R-FCN), and region-based Fast R-CNN are popular two-stage methods. The second strategy uses regression-based one-stage detection techniques that estimate the borders and directly distinguish particular groups. Although Although these one-stage techniques provide a higher processing speed than the two-stage method, their accuracy is typically a little worse. Among the renowned algorithms in this field are RetinaNet, The YOLO series and Single Shot MultiBox Detector (SSD) [14].

Although there is a processing bottleneck, the double detection method has certain detection accuracy improvements over the single detection method. The single detection method has a stronger detection effect even though its accuracy is a little lower. Still, the YOLO network has issues, including a lengthy training period and trouble differentiating certain targets over several scenarios, despite being An illustration of a single detection network. The efficiency and accuracy will increase if single detection techniques like YOLO can be modified, and non-target interference mistakes in the identification can be successfully controlled by certain approaches [15].

The discipline of computer vision has become more interested in the YOLO collection of algorithms. Because it maintains a small model size while maintaining a high degree of realism, YOLO is quite popular. Because one GPU can be used to train YOLO models., a broad They can be used by a variety of developers. Experts in ML can deploy it at a fair price on edge hardware or in the cloud. YOLOv8, the latest and most advanced YOLO technique, may be used for applications such as image classification, object recognition, and segmentation. Ultralytics, who also created the important YOLOv5 model that shaped the industry, produced Yolo v8. YOLOv8 features some architectural improvements and updates over YOLOv5 [16]. Arsitektur YOLOv8 can be seen in Figure 2.

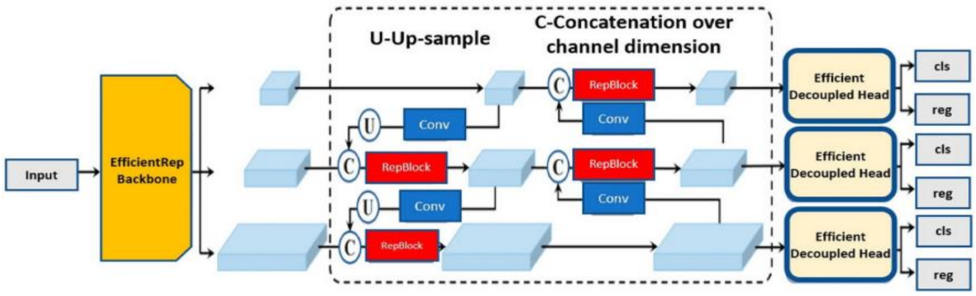


Figure 2. Arsitektur YOLOv8

YOLO has an advantage over other algorithms because its process is fast and efficient in real time and has a fairly good accuracy. YOLO has several versions and YOLOv8 is the 2023 version, which is faster, more accurate, and more effective than the previous version [17]. Table 1 illustrates the architecture of YOLOv8.

YOLOv8 merges features extracted from various resolution feature maps to generate multi-scale features. The backbone takes the input image and down samples it five times, producing five scale feature maps, each represented by the notation {P1, P2, P3, P4, P5}. Extracting these relevant elements from the input image is the main purpose of the backbone. The neck, which serves as a link between the detecting head and the extracted characteristics, comes after the backbone . In order for the detection head to use the most useful representation for object detection, the neck smooths out these characteristics and makes it easier for them to fuse. Lastly, object detection responsibilities will be handled by the detection head. It's important to remember that in YOLOv8, the scale set {P3, P4, P5} plays a crucial role, with each scale specifically responsible for detecting objects of a particular size range – small objects for P3, medium objects for P4, and large objects for P5. This division of labor across scales strengthens the model's overall detection accuracy [18].

Table 1. YOLO Object detection Models Timeline

Model	Speed	Accuracy	Year
YOLOv1	Medium	Medium	2015
YOLOv2	Fast	High	2016

YOLOv3	Very Fast	High	2018
YOLOv4-5	Very Fast	Very Fast	2020
YOLOv6-7	Very Fast	Very Fast	2022
YOLOv8	Very Fast	Very Fast	2023

2.4. Object Detection

Object detection is a core problem in computer vision. Detection pipelines generally start by extracting a set of robust features from input images (Haar, SIFT, HOG, convolutional features). Then, classifiers or To locate items in the feature space, localizers are utilized. These localizers or classifiers are applied to a portion of the picture or all of the picture in a sliding window. We illustrate the main parallels and divergences between the YOLO detection method and a number of leading detection frameworks [19].

In the domains of industrial automation, autonomous vehicles, computer vision, and other, object detection is crucial. Detecting objects in real time is a challenging process. Traditional target detection is inferior to deep learning in object detection. Region proposal object detection algorithms are one type of deep learning technique that creates region proposal networks and then classifies them. These include SPPnet, FastRCNN, Faster-RCNN, Region-based Convolutional Neural Networks, and others. Regression object detection methods like SSD and YOLO are used to simultaneously create and classify region proposal networks. The several real-time object identification techniques based on YOLO are compiled in this study [20].

An application needs a frame processing speed of several hundred milliseconds for object detection in real time. In other words, depending on the needs of the application, It is sufficient to have real-time object detection speeds of roughly 3–5 or 10 frames per second. But in the case of the current YOLO, there could be a major issue with real-time processing if the frame rate transmitted by the camera is greater than the object recognition service rate [21].

2.5. Confusion Matrix

A confusion matrix, also known as an error matrix, compares the actual classification results with the classification results that the system (model) produced. The Confusion Matrix, which takes the shape explains how well the categorization A set of test data with known actual values is used to test the model. in a matrix table. Four terms in the confusion matrix represent the results of the categorization process: False Negative (FN), False Positive (FP), True Positive (TP), and True Negative (TN). The four terms in the Confusion Matrix table are explained is provided below: A TP is a positive result that the system (model) accurately predicts. One instance is when an object is identified as An individual object created by the system (model). Negative data that the system (model) accurately predicts is known as TN. For instance, the system (model) may be aware of an item but fail to recognize it as such. The term FP, is negative data that the system (model) predicts will be positive. For instance, when an object is present, its status as a person object is acknowledged by the system (model). d. Positive results known as FN are those that the system (model) predicts to be negative. For instance, when an object is present, the system (model) recognizes that it is an object [22].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN+FT} \tag{4}$$

$$\text{F1-Score} = 2x\,\frac{PRECISION \times RECALL}{PRECISION + RECALL} \tag{5}$$

The confusion matrix reveals the classifier's performance, what it is getting right, and the many kinds of mistakes the classifier could make. The confusion matrix's measurements aid in selecting the optimal approach to improve the model's performance. Confusion matrices are employed in supervised learning techniques because they may be created for datasets with known target/output values [23].

A ML The confusion matrix is used to evaluate the model's performance. The confusion matrix is one such matrix. displays both the actual and expected categorization predictions [24].

3. RESULTS AND ANALYSIS

This system is designed to detect violations of student ethics in the classroom, especially eating and drinking activities, using This Raspberry Pi device uses YOLO technology. The Raspberry Pi is linked to the camera to capture real-time images, which are then processed by the YOLO model to detect inappropriate activities.

3.1. Dataset Collection

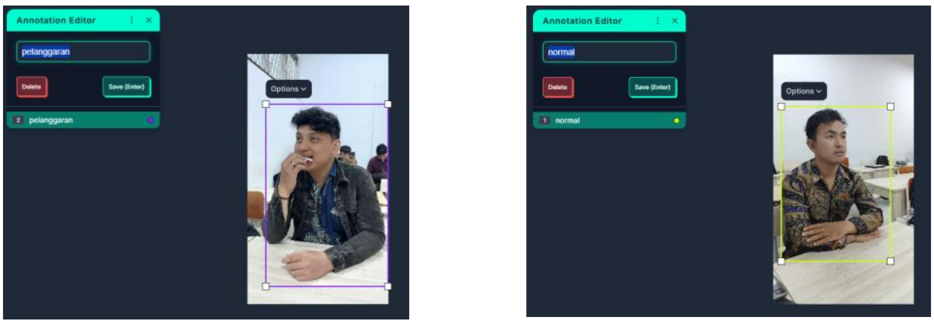
Dataset collection performs data training to get a model with good accuracy from the collected student activity dataset. The dataset was collected manually and through the Roboflow platform, consisting of a total of 450 images (250 images of violations: eating/drinking and 200 normal images). Figure 4 is the student activity dataset.

Figure 4. Student Activity Dataset

Each image was annotated using the YOLO format via Roboflow, which includes bounding box coordinates and the classes "violation" and "normal." The collection process was conducted in a classroom environment with various scenarios, viewpoints, and lighting conditions to ensure representative data. This dataset contains student activities in the classroom with violation behavior (eating and drinking) and normal behavior.

3.2. Dataset Notation

Each of the millions of distinct photos in Strong deep learning image classifiers are trained using the ImageNet dataset for general use, has been annotated to explain the items it includes. Although they are often smaller, hundreds to thousands of annotated photos are commonly observed in datasets that are used to train robust classifiers for medical images. Many people believe that one of the biggest barriers to deep learning system development is the work necessary to curate these training datasets [25]. Figure 4 shows the annotation violation dataset and the normal annotation dataset. The image annotation process uses roboflow. This labeling uses annotations in the YOLO annotation format.



(a) (b)

Figure 4. (A) Annotation Offense Dataset (B) Normal Dataset Annotation

3.3. Pre Processing

First, A grid is created from the input image of SxS frames (where S is a predefined number multiple of 32) as part of the visual processing. YOLO forecasts the coordinates of the bounding box (x-center, y-center, width, height), the likelihood that it includes an object, and the likelihood of a particular class for every frame and every potential associated anchor box. Class predictions and confidence thresholds eliminate unreliable or irrelevant detections. Based on the overlap and confidence of the detections, Absence of maximum suppression removes duplicate detections to lessen detection redundancy. A series of enclosing boxes with the class labels and corresponding confidence probabilities is the end result.[26]. This process helps the model become robust to various conditions.

Figure 5. Pre Processing Dataset

In Figure 5, the preprocessing stage aims to improve image quality, making it easier for the character detection and recognition process.

3.4. Training Dataset

Labeled picture data preparation, model parameter configuration, and weight optimization via iterative learning are all part of the YOLOv8 training phase. There are training and validation sets inside the dataset, with annotations following the YOLO format: Class ID and normalized bounding box coordinates.

3.4.1. Epoch Value Testing, Batch, and YOLOv8 Models

In Figure 6 are the YOLOv8 model's performance measures after 55 epochs of training. The training metrics are displayed in the top row, while the bottom row shows the corresponding validation metrics.

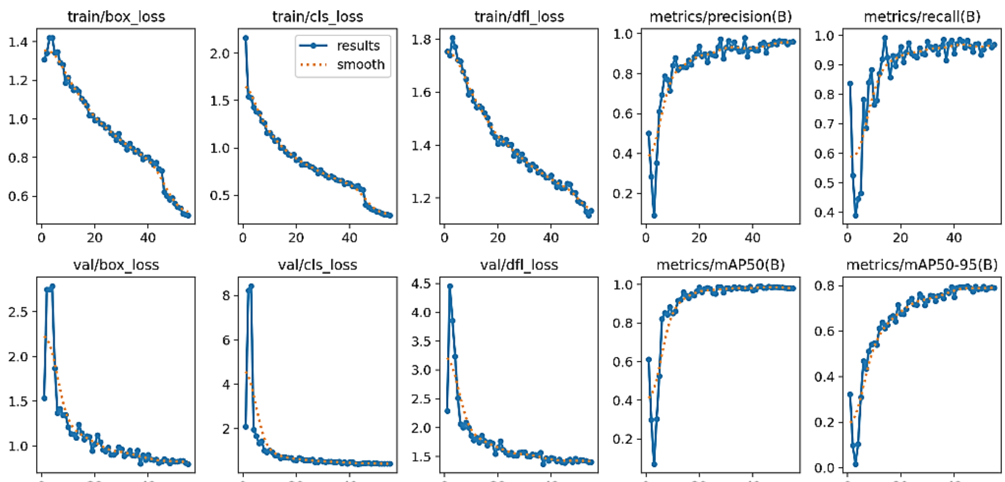


Figure 6. YOLOv8 Results Performance Graph

The consistent decrease in loss, as well as the increase in precision, recall, and mAP on the validation data, indicate that the model can identify and categorize items well without overfitting.

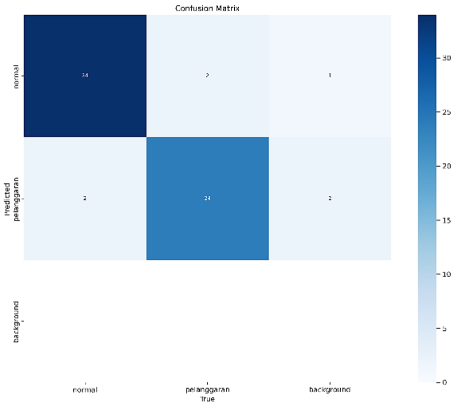


Figure 7. Confusion Matrix

The Confusion Matrix image that has been tested, which includes True Positive, True Negative, False Positive, and False Negative are abbreviated as TP, TN, FP, and FN, is depicted in the Figure 7. Each class is classified into multiple classes using the one-vs-all method in the Table 2.

Table 2. Confusion Matrix

Kelas	TP	TN	FP	FN
Normal	34	24	2	4
Pelanggaran	24	34	3	5

In addition, distribution focal loss (DFL) is used to improve the accuracy of bounding box prediction. The train/dfl_loss value is initially around 1.8 and drops consistently to close to 1.0. In the validation data, the initial value of about 4.0 drops to about 1.2. This decrease shows that the model is able to perform spatial bounding box prediction more precisely.

Precision increases rapidly from an initial value of 0.0 to close to 1.0 at the 20th epoch and then stabilizes. This indicates that the proportion of correct object detections gets higher as the epochs increase. Recall also increased from around 0.2 to above 0.9, indicating that The majority of the objects in the validation image could be detected by the model.. The significant increase in the first few epochs indicates the efficiency of the training process.

Next, use the mean Average Precision (mAP) measure to assess the model's performance, mAP@0.5 (metrics/mAP50(B)): Increased from 0.0 to around 0.95, indicating very high performance while detecting objects mAP@0.5:0.95 at a threshold of 0.5 IoU (metrics/mAP50-95(B)).

3.5. Testing Dataset

The goal of the YOLOv8 testing procedure is to assess the performance of the model. The evaluation is done utilizing common measures like recall, accuracy, and mAP. The results of these tests become the basis for concluding whether the model is feasible to be applied to real application environments.

In Figure 8, the results show that the performance is good enough to detect the system of ethical violations of students who are eating and drinking and the normal state of students. The test results show that the system is capable of detecting students' eating and drinking behavior in real-time with high accuracy. With a mAP@0.5 of 95% and an F1-score close to 0.93, the system is highly reliable under normal lighting conditions. These results also indicate that YOLOv8 outperforms YOLOv5, as found by Sary et al. (2023), who reported an accuracy of 89% in a similar scenario. This research also outperforms the YOLOv4 model in public CCTV behavior detection systems, as in the study by Magalhães et al. (2021), which faced challenges in distinguishing small objects. Thus, this research has novelty in combining edge technology (Raspberry Pi), real-time detection methods, and academic context. Its practical implementation has great potential for application in various educational institutions.

The test results show that the system is capable of detecting students' eating and drinking behavior in real-time with high accuracy. With a mAP@0.5 of 95% and an F1-score close to 0.93, the system is highly reliable under normal lighting conditions. These results also indicate that YOLOv8 outperforms YOLOv5, as found by Sary et al. (2023), who reported an accuracy of 89% in a similar scenario.

This research also outperforms the YOLOv4 model in public CCTV behavior detection systems, as in the study by Magalhães et al. (2021), which faced challenges in distinguishing small objects. Thus, this research has novelty in combining edge technology (Raspberry Pi), real-time detection methods, and academic context. Its practical implementation has great potential for application in various educational institutions.

Although the overall system performance is already quite good, it must be acknowledged that there are still cases where the system makes detection errors. These can occur in the form of false positives, where normal behavior is mistakenly interpreted as a violation, or false negatives, where clear violations are not detected due to insufficient dataset. Additionally, in some scenarios, the bounding box may not appear even though there is an object that should be detected. This phenomenon can be caused by various factors, such as extreme lighting variations, partial or complete object occlusion, high similarity between the target object and the background, or even a confidence threshold that is too high, causing detections with low confidence scores to be ignored. These challenges underscore the importance of continuous testing under various environmental conditions and the potential for system refinement or hyperparameter adjustment for future research.

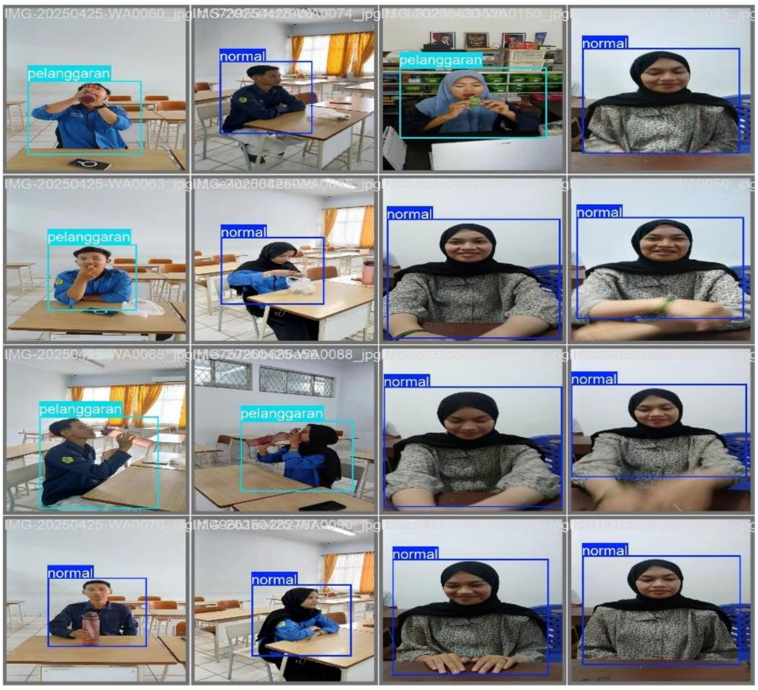


Figure 8. Training Results

4. CONCLUSION

This study successfully designed a system for monitoring student behavior in the classroom using the YOLOv8 object detection method. This system effectively detects eating and drinking behavior as violations of student ethics in real time. The manually collected dataset was successfully developed and produced a model with fairly good performance, showing stable results, marked by decreasing loss values and increasing evaluation parameters, such as mAP, recall, and precision. The results indicate that the trained model is highly effective, achieving a 95% mAP@0.5 and 75% mAP@0.5:0.95. Test results prove that the system can work stably under good lighting and certain viewing angles. This research makes a significant contribution to the field of automatic and real-time academic ethics monitoring. In the future, this system can be further developed with improvements in multi-behavior detection, low-light processing, and integration with an automatic warning system.

Real-world testing shows that the system can recognize student behavior in the classroom, particularly eating and drinking, in real-time with a reasonably high level of confidence under specific lighting conditions and viewing angles. This demonstrates that the YOLOv8-based approach is effective in automatically detecting student ethical violations and has the potential for widespread application in technology-based learning environment monitoring systems. Although the system functions well, there are still limitations in terms of dependence on environmental conditions and the scope of behaviors that can be detected. Therefore, further development is recommended to enhance the model's robustness under varying lighting conditions and camera positions, expand the types of behaviors recognized, utilize hardware with higher computational capabilities, and integrate an automatic notification system for more responsive and adaptive monitoring in real-world environments.

REFERENCES

[1] U. Handalage, L. Kuganandamurthy, and U. Hdi, “Real-Time Object Detection Using YOLO: A Review SEE PROFILE Real-Time Object Detection using YOLO: A review,” no. May, 2021, doi: 10.13140/RG.2.2.24367.66723.

[2] G. Bannerjee, U. Sarkar, S. Das, and ..., “Artificial intelligence in agriculture: A literature survey,” Int. J. …, no. June, 2018, [Online]. Available: https://www.researchgate.net/profile/Gouravmoy-Banerjee/publication/326057794_Artificial_Intelligence_in_Agriculture_A_Literature_Survey/links/5b35ab970f7e9b0df5d83ec6/Artificial-Intelligence-in-Agriculture-A-Literature-Survey.pdf

[3] D. J. Shin and J. J. Kim, “A Deep Learning Framework Performance Evaluation to Use YOLO in Nvidia Jetson Platform,” Appl. Sci., vol. 12, no. 8, 2022, doi: 10.3390/app12083734.

[4] Y. Li, J. Wang, J. Huang, and Y. Li, “Research on Deep Learning Automatic Vehicle Recognition Algorithm Based on RES-YOLO Model,” Sensors, vol. 22, no. 10, 2022, doi: 10.3390/s22103783.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 2020, [Online]. Available: http://arxiv.org/abs/2004.10934

[7] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS,” Mach. Learn. Knowl. Extr., vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.

[8] Y. Cao, Q. Cao, C. Qian, and D. Chen, “YOLO-AMM: A Real-Time Classroom Behavior Detection Algorithm Based on Multi-Dimensional Feature Optimization,” Sensors, vol. 25, no. 4, pp. 1–21, 2025, doi: 10.3390/s25041142.

[9] F. Yang, T. Wang, and X. Wang, “Student Classroom Behavior Detection Based on YOLOv7+BRA and Multi-model Fusion,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 14357 LNCS, pp. 41–52, 2023, doi: 10.1007/978-3-031-46311-2_4.

[10] V. Viswanatha, R. K. Chandana, and A. C. Ramachandra, “IoT Based Smart Mirror Using Raspberry Pi 4 and YOLO Algorithm: A Novel Framework for Interactive Display,” Indian J. Sci. Technol., vol. 15, no. 39, pp. 2011–2020, 2022, doi: 10.17485/ijst/v15i39.1627.

[11] S. Das, A. Dey, A. Pal, and N. Roy, “Applications of Artificial Intelligence in Machine Learning: Review and Prospect,” Int. J. Comput. Appl., vol. 115, no. 9, pp. 31–41, 2015, doi: 10.5120/20182-2402.

[12] K. C. A. Khanzode and R. D. Sarode, “Advantages and Disadvantages of Artificial Intelligence and Machine Learning: A Literature Review,” Int. J. Libr. Inf. Sci., vol. 9, no. 1, pp. 9–10, 2020, [Online]. Available: http://www.iaeme.com/IJLIS/index.asp30http://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1JournalImpactFactor%0Awww.jifactor.comhttp://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1

[13] A. Kish, “Machine Learning: A Review of Methods and Applications,” Researchgate.Net, vol. 10, no. 07, 2018, [Online]. Available: https://www.researchgate.net/profile/Adam-Kish-2/publication/327645425_Machine_Learning_A_Review_of_Methods_and_Applications/links/5b9b679a299bf13e602d5bcd/Machine-Learning-A-Review-of-Methods-and-Applications.pdf

[14] X. Wang, H. Gao, Z. Jia, and Z. Li, “BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8,” Sensors (Basel)., vol. 23, no. 20, 2023, doi: 10.3390/s23208361.

[15] S. A. Magalhães et al., “Evaluating the single-shot multibox detector and yolo deep learning models for the detection of tomatoes in a greenhouse,” Sensors, vol. 21, no. 10, pp. 1–24, 2021, doi: 10.3390/s21103569.

[16] F. M. Talaat and H. ZainEldin, “An improved fire detection approach based on YOLO-v8 for smart cities,” Neural Comput. Appl., vol. 35, no. 28, pp. 20939–20954, 2023, doi: 10.1007/s00521-023-08809-1.

[17] A. A. Rasjid, B. Rahmat, and A. N. Sihananto, “Implementasi YOLOv8 Pada Robot Deteksi Objek,” J. Technol. Syst. Inf., vol. 1, no. 3, p. 9, 2024, doi: 10.47134/jtsi.v1i3.2969.

[18] Z. J. Khow, Y. F. Tan, H. A. Karim, and H. A. A. Rashid, “Improved YOLOv8 Model for a Comprehensive Approach to Object Detection and Distance Estimation,” IEEE Access, vol. 12, no. April, pp. 63754–63767, 2024, doi: 10.1109/ACCESS.2024.3396224.

[19] G. Plastiras, C. Kyrkou, and T. Theocharides, “Efficient convnet-based object detection for unmanned aerial vehicles by selective tile processing,” ACM Int. Conf. Proceeding Ser., 2018, doi: 10.1145/3243394.3243692.

[20] B. S. Rekha, A. Marium, G. N. Srinivasan, and S. A. Shetty, “Literature Survey on Object Detection using YOLO,” Int. Res. J. Eng. Technol., vol. 07, no. 06, pp. 3082–3088, 2020, [Online]. Available: https://www.irjet.net/archives/V7/i6/IRJET-V7I6576.pdf

[21] J. Lee and K. il Hwang, “YOLO with adaptive frame control for real-time object detection applications,” Multimed. Tools Appl., vol. 81, no. 25, pp. 36375–36396, 2022, doi: 10.1007/s11042-021-11480-0.

[22] A. Amperawan, D. Andika, M. Anisah, S. Rasyad, and P. Handayani, Confusion Matrix Using Yolo V3-Tiny on Quadruped Robot Based Raspberry PI 3B + . Atlantis Press International BV, 2024. doi: 10.2991/978-94-6463-386-3_56.

[23] S. Sathyanarayanan and B. R. Tantri, “Confusion Matrix-Based Performance Evaluation Metrics,” no. December,

2024, doi: 10.53555/AJBR.v27i4S.4345.
[24] I. P. Sary, S. Andromeda, and E. U. Armin, “Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images,” Ultim. Comput. J. Sist. Komput., vol. 15, no. 1, pp. 8–13, 2023, doi: 10.31937/sk.v15i1.3204.
[25] K. A. Philbrick et al., “RIL-Contour: a Medical Imaging Dataset Annotation Tool for and with Deep Learning,” J. Digit. Imaging, vol. 32, no. 4, pp. 571–581, 2019, doi: 10.1007/s10278-019-00232-0.
[26] D. Valencia, E. Munoz Espana, and M. Munoz Anasco, “Impact of the preprocessing stage on the performance of offline automatic vehicle counting using YOLO,” IEEE Lat. Am. Trans. , vol. 22, no. 9, pp. 723–732, 2024, doi: 10.1109/TLA.2024.10669248.

BIBLIOGRAPHY OF AUTHORS

Ifroh Intan Sepulau was born in Palembang, South Sumatra, Indonesia. She obtained her Bachelor of Applied Electrical Engineering degree from the D4 Electrical Engineering Study Program, Department of Electrical Engineering, Sriwijaya State Polytechnic. Her research interests include computer vision, deep learning, edge computing-based object detection systems, and embedded system development.

RD.Kusumanto, is a lecturer of Electrical Engineering, Departement of Electrical Engineering, at Sriwijaya State Polytechnic, Indonesia

Nyayu Latifah Husni, is a lecturer of Electrical Engineering, Departement of Electrical Engineering, at Sriwijaya State Polytechnic, Indonesia.