Journal of Information Systems and Informatics
Vol. 6, No. 4, December 2024 e-ISSN: **2656-4882** p-ISSN: **2656-5935**
DOI: 10.51519/journalisi.v6i4.944 Published By DRPM-UBD

Evaluating YOLOv5 and YOLOv8: Advancements in Human Detection

Joel Hamim Sim[1], Ni'matul Ma'muriyah[2], Andik Yulianto[3]

[1,2,3]Information Systems, Faculty of Computer Science, Batam International University, Batam, Indonesia
Email: [1]2131145.joel@uib.edu, [2]nimatul@uib.ac.id, [3]andik@uib.ac.id

Abstract

The YOLO (You Only Look Once) method is a state-of-the-art approach in real- time object detection, known for its high-speed image processing capabilities. RecentlyYOLO versions have differed in performance, particularly in terms of detection accuracy and computational efficiency. The objective of this study is to assess the effectiveness and performance of YOLOv5 and YOLOv8 in real-time human detection applications using the SEMMA (Sample, Explore, Modify, Model, and Assess) methodology also. The dataset was processed through the Roboflow platform, which facilitated both the dataset management and the labeling process. Roboflow's tools streamlined the annotation of images, ensuring consistent labeling for deep learning model training and evaluation. F1 score, recall score, and precision score are compared both YOLOv5 and YOLOv8 to evaluate the performance of these architectures. The result of the evaluations shows that the performance of the YOLOv8 is better than the YOLOv5 which, YOLOv5 achieved F1-score equal 0.5865 (58%), recall score equal 0.83 (83%), and precision score of 0.4535 (45%). Meanwhile, YOLOv8 demonstrated better performance, with F1-score of 0.7921 (79%), recall score of 0.8289 (82%), and precision score of 0.7585 (75%). Base on the evaluations, we concluded that the performance of the YOLOv8 model is greater than the YOLOv5 model for Precision, and F1-Score, while YOLOv5 has slightly better score on recall. The contribution of this study is going to implemented into Audio guidance for the blind's prototype that have been developing in previous study.

Keywords: YOLOv5, YOLOv8, Human Detection, Roboflow, F1-score, Precision, Recall

1. INTRODUCTION

The Sustainable Development Goals (SDGs) aim to enhance global welfare, with the 9th goal emphasizing the importance of innovation, industry, and infrastructure. A key aspect of this goal is encouraging research that advances technology for industrial applications. Human detection using machine learning algorithms is one such technological implementation, widely applied in industries such as security, attendance systems, and surveillance. For instance, facial recognition-based attendance systems have demonstrated the practical value of these algorithms in industrial settings [1]. In previous study used Local Binary Pattern Histogram

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

(LBPH) algorithm and Convolution Neural Network (CNN) to detect human emotion and facial [2].

With the development of machine learning algorithms, the process of object detection or human detection or face detection is easier to do with a good level of accuracy. The 3 machine learning algorithms that have been widely used in the research in the last 5 years are Histogram of Oriented Gradients (HOG), Mask R-CNN, and You Only Look Once (YOLO). The following are some previous studies that use the HOG algorithm in detecting objects such as mango fruits[3] the following are human detection studies with the HOG algorithm [4], [5], [6]. The following is a summary of the advantages and disadvantages of HOG in object detection or human detection, the HOG algorithm is effective in detecting objects based on visual features, with satisfactory accuracy in stable lighting conditions and simple environments. This method utilizes gradient orientation patterns to effectively recognize shapes and edges, but HOG has some limitations, namely HOG performance is very sensitive to lighting and shade variations.

The region-based convolution neural network (R-CNN) is a human detection algorithm based on a convolution neural network. The object detection process in this algorithm uses 3 modules, namely the first module is a proposal for the generation of categories from independent regions, the second module is feature extraction using the convolution neural network method, and the last module is the evaluation stage of object detection or classifying objects using the PASCAL VOC 2010-12 and ILSVRC2013 algorithm [7]. Object detection with R-CNN takes 10 to 45 minutes for one image, therefore some researchers have made improvements focusing on reducing the processing time, the following are improvements to the R-CNN algorithm to Fast R-CNN or Faster R-CNN, and Mask R-CNN. The Mask R-CNN algorithm used to detect humans can be found in the following studies[8], [9]. Based on the results of previous research, the R-CNN algorithm has a shortcoming in that the process time is quite long and for the determination of detection results, a new algorithm is needed as an evaluator and determinant of object detection or classification object. In a study that compares the Mask R-CNN algorithm with the YOLO algorithm in object detection, the results were obtained that the YOLO algorithm is better than the Mask R-CNN, based on the results of the study, YOLO is better in a. object detection ability; and b. faster computation times compared to the Mask R-CNN algorithm [10].

The next object or human or face detection algorithm is YOLO, according to research [10], the YOLO algorithm has the advantage of speed in processing object detection with good accuracy so that the YOLO algorithm is very suitable for object detection that requires real-time time. The need for object detection or real-time human detection is greatly increasing, especially for security needs. The following are some human detection studies using the YOLO algorithm [11] [12] [13]. The use of the YOLO algorithm is not only used as human detection, but can also be applied

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

to vehicle counting systems or determine the type of vehicle as has been done by the following research[14], [15]. Implementation of YOLOv8 in research focusing on human activity detection using YOLOv8 deep learning techniques [16]. The results showed that training of human activity photos using YOLOv8 produced more accurate and precise results. However, the limitation of this study is that YOLOv8 requires a fairly long training time, reflecting the increasing complexity of the model. Meanwhile, the Yolov5 study[17] presented the results of facial recognition photo training using YOLOv5, with a Precision score of 0.253, Recall 0.247, and F1-Score 0.25. The weaknesses of this study include a reliance on normal lighting conditions, limited variation in the training dataset, and a lack of generalization to real-world data. YOLOv5 shows a drop in performance in extreme lighting conditions, obstructed faces, or when detecting multiple individuals in a video. In addition, the small size of the dataset and the limited test environment make it challenging to apply these results to real-world scenarios.

The results of the literature review that the team has carried out, in the last 5 years the development of YOLO has been very fast, various types of YOLO that have been developed by previous researchers, namely YOLOv2, YOLOv3, YOLOv4, YOLOv5 and YOLOv8. Each type has advantages compared to the previous type. To make it easier for researchers to determine which type of YOLO is better used in the object detection process or human detection, in this study the team conducted a comparative study between YOLOv5 and YOLOv8. The main purpose of this study is to determine between the two types that have good performance and accuracy or efficiency.

In previous study was developed prototype an audio guidance for the blind by using raspberry pi and fuzzy logic controller and to navigated the blind[18]. Our prototype has weakness in real time object recognition or human recognition, so we hope with this study can improve our prototype made better performance. The presentation of writing this article in accordance with the template of this journal begins with a background explanation that explains the phenomena that support the research team to take this research topic, then continues with an explanation of the method used in making comparisons between the 2 types of YOLO, starting with a general explanation of YOLO, YOLOv-5, and YOLOv8, the dataset used in this study, and evaluation metrics. Then an explanation of the summary of the results of the literature review which is used as the main reference of the research team. The research model will be discussed in the next sub-chapter, which explains the YOLO architecture and also the flowchart of the object detection process or human detection. The next chapter is Result and Discussion in this section will discuss more about the results of research that has been carried out and the analysis of the results of the research, and the last chapter is a summary of the results of the research that has been carried out and the conclusion of the results obtained.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2. METHODS

2.1 YOLO

You Only Look Once (YOLO) is an algorithm developed for real-time object detection [19]. The detection system works by repurposing classifiers or localizers to perform object detection. A model is applied to an image at multiple locations and scales. The region of the image that receives the highest score is considered a detection. YOLO uses an artificial neural network (ANN) approach to detect objects in an image. The network divides the image into several regions and predicts the bounding boxes and probabilities for each region. These bounding boxes are then compared to each predicted probability. YOLO has several advantages over classifier-based systems, as it considers the entire image during testing, providing predictions informed globally across the image. This also makes its predictions, using neural network synthesis, significantly faster than Region-Convolutional Neural Networks (R-CNN), which require thousands of applications to an image, making YOLO several times faster than R-CNN.

2.2 YOLOv5

The standard architecture of YOLOv5 has evolved into a three-part structure: backbone, neck, and head [20]. The backbone, Darknet 53, is a network architecture designed for feature extraction, utilizing small filter windows and residual connections. Acting as a bridge between the backbone and head, the neck gathers and refines features extracted from the backbone, enhancing both spatial and semantic information across multiple scales. Finally, the head contains three branches, each responsible for predicting features at different scales, producing bounding boxes, class probabilities, and confidence scores. The model then applies Non-maximum Suppression (NMS) to filter overlapping bounding boxes.

2.3 YOLOv8

YOLOv8 is the latest version in the YOLO object detection model series, following YOLOv5. It brings advancements with a redesigned neural network architecture [12]. This version incorporates two neural networks: the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN). Additionally, YOLOv8 introduces a new labeling tool that streamlines the annotation process, offering features like automatic labeling, shortcut- based labeling, and customizable hotkeys. These enhancements make annotating images for model training faster and more efficient.

2.4 Datasets

The datasets utilized in this study comprise Roboflow, roboflow widely accessible to

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the public and extensively employed in research and academia. These datasets serve as valuable resources for training and evaluating the performance of machine learning models, ensuring a comprehensive exploration of the capabilities and limitations of YOLO and its iterations within the context of image processing.

2.5 Evaluation Metrics

A confusion matrix is a tool for evaluating the performance of a machine learning model, displaying the relationship between actual classifications and predicted classifications. It includes four possible outcomes: True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP), each reflecting combinations of actual and predicted values. Table 4 outlines these definitions: TP (True Positive) is the count of correctly classified positive samples; TN (True Negative) is the count of correctly classified negative samples; FP (False Positive) is the count of negative samples incorrectly classified as positive; and FN (False Negative) is the count of positive samples incorrectly classified as negative. Metrics such as precision, recall, and F1-score, derived from the confusion matrix, provide insight into the model's performance. An illustration of the confusion matrix can be seen in Figure 1.

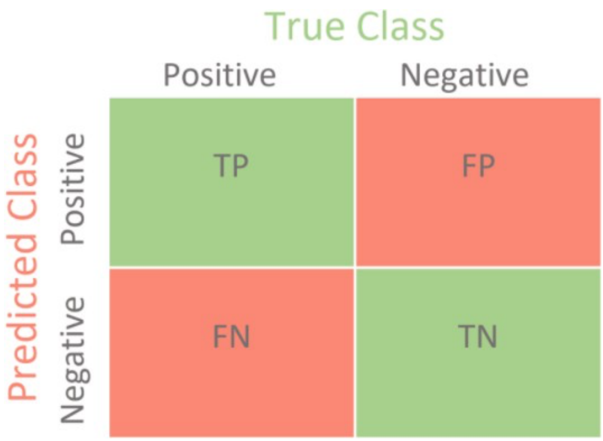

Figure 1. Confusion Matrix

Precision is the ratio of TP to the total number of predicted positive data. In the denominator, there is the variable FP as the divisor. This can be written using Equation 1

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

On the other hand, recall is defined as the ratio of TP to the total number of actually positive instances. The denominator includes FN as the divisor, and it can be written using Equation 2

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

When recall is very high, precision will be very low, and vice versa. There is a trade-off relationship between precision and recall. This trade-off relationship implies that the sum of these two variables equals 1. The harmonization of the average between precision and recall is called the F1-score. Based on Equation 3, the best value for the F1-score is 1.0, while the worst value is 0.0.

$$F1 = \frac{2 \times precision \times recall}{precision+recall} \tag{3}$$

2.6 SEMMA

To support YOLOv5 and YOLOv8 to detect human we use The SEMMA Method. This methodology approach commonly applied in several research in data mining area such as prediction, clustering or classification[21][22]. Figure 2 shows framework of the SEMMA method. This method is structured into several stages:

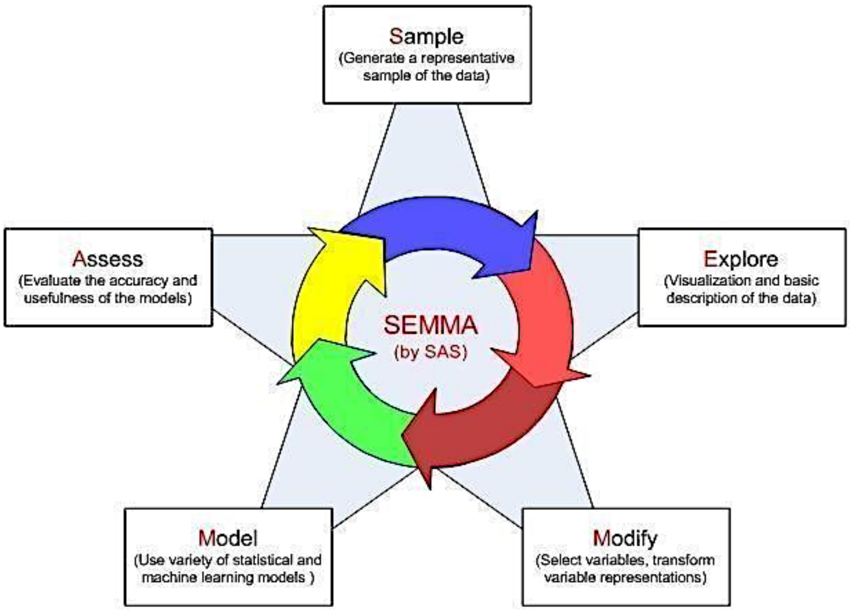

Figure 2. Framework of The SEMMA Method [23]

1) **Sample (Sampling)**: Identification and collection of relevant sample data for human detection objectives. This involves selecting datasets that align with the research scope.
2) **Explore (Exploration)**: Initial analysis of sample data to understand its

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

characteristics. Exploration includes data visualization, pattern identification, and further understanding of the dataset's nature.

3) **Modify (Modification)**: Data preprocessing to enhance quality and comprehensibility. Modifications may involve normalization, filling missing data, or other transformations necessary to prepare the data for further processing.
4) **Model (Modeling)**: Development and training of the model for human detection. Model selection, such as YOLO v5 or YOLO v8, and other configurations are done in this stage to achieve optimal results.
5) **Assess (Assessment)**: Evaluation of the model's performance against the testing dataset. Accuracy, precision, and recall measurements can be used to gauge the effectiveness of human detection and ensure that the model provides reliable results.

Previous study which applying the SEMMA method tailored for human detection, this research can provide a comprehensive understanding regarding the implementation of YOLO models in this context [24].

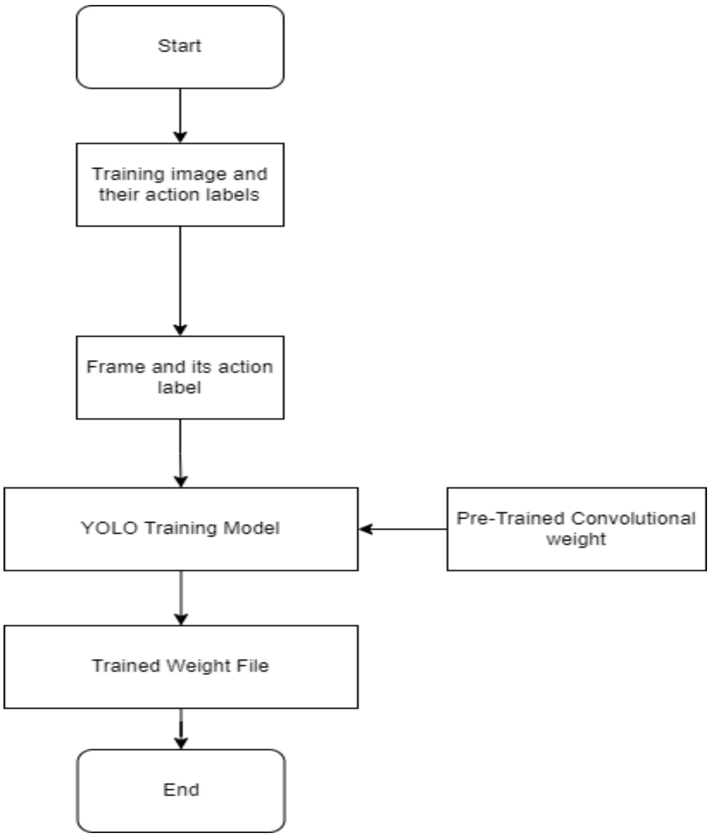

Figure 3. Flowchart for phase of Human Detection

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In this study experiment model using The Architecture of YOLO have 24 convolutional layers and 2 fully connected layers.

1) The total number of action classes.
2) A text file containing paths to all the frames to be used for training.
3) A text file listing the names of all action classes.
4) The directory path where the trained weight files will be saved.
5) A configuration file describing the YOLO architecture layers, as detailed in Figure 3.
6) Pre-trained convolutional weight files.
7) It is important to note that the number of filters in the second-last layer of YOLO's configuration file (.cfg file) is not chosen arbitrarily; it is determined by the total number of action classes.

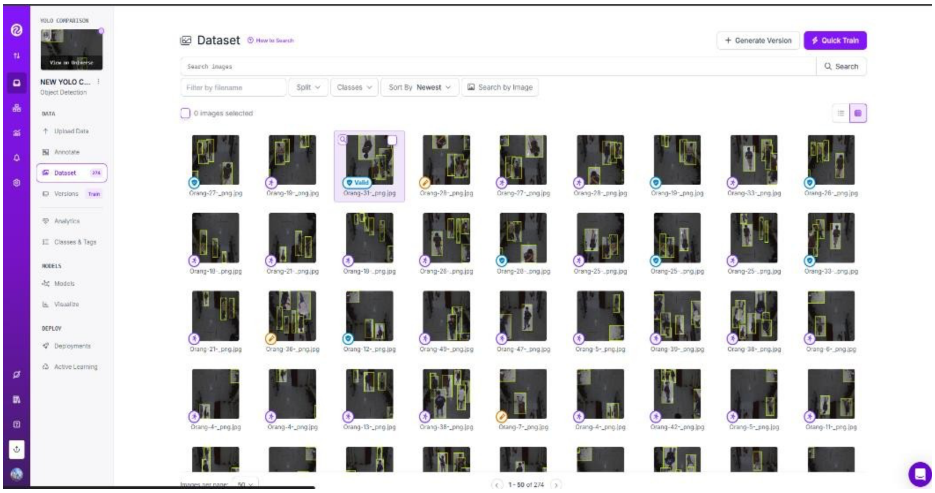

Figure 4. Sample image Roboflow

Figure 4. shows a dataset interface from an object detection platform. It features images annotated with bounding boxes that emphasize identified objects, particularly people. The annotations are rendered in yellow, signifying the objects classified under the "person" category. This dataset comprises a total of 274 images, which are categorized by various statuses, such as validation, represented by icons including checkmarks, walking symbols, and other indicators. The platform offers capabilities for managing datasets, training, and deploying models, as indicated by options such as "Quick Train" and "Generate Version" available in the interface.

3. RESULTS AND DISCUSSION

3.1. Dataset Training

Dataset training was carried out on Google Colab by dividing the dataset into

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

16 batches. Training iterations are carried out for 100 epochs using the yolov5s.yaml configuration, which is an important file that organizes various aspects of the YOLOv5s and YOLOv8s models with the backbone model used being 's' or small. Figure 6 is a training process that has been carried out as many as 100 epochs. From the training, a weight named 'best.pt' was obtained which was used as a weight in the calculation system. Dataset training process can be seen on Figure 5.

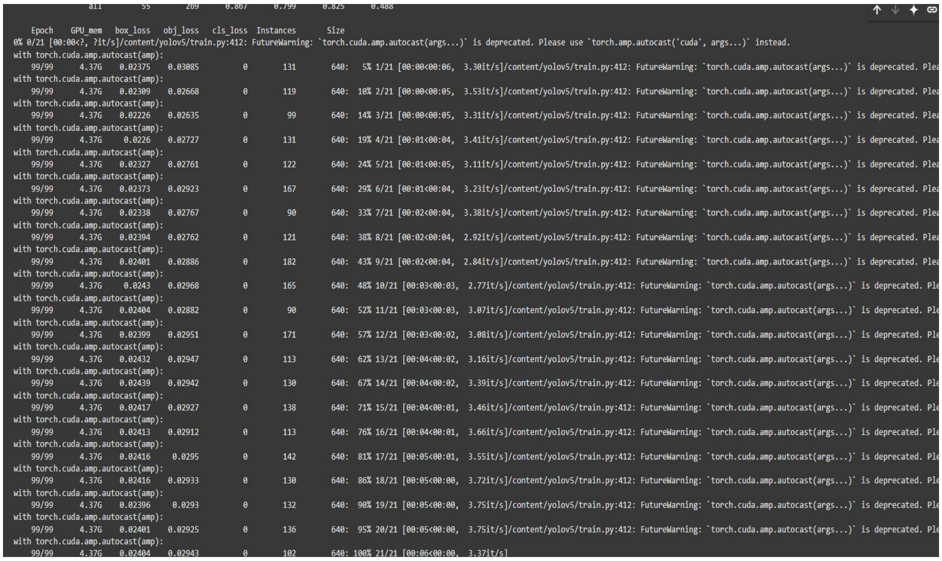

Figure 5. Dataset Training Process

3.2. Experimental Results

The performance results of YOLOv5 and YOLOv8 can be seen in Figure 6 and Figure 7, also the human detection result can be seen on Figure 8 and Figure 9. Figure 6 is the performance result of YOLOv5, the image represented the evolution of losses and evaluation metrics throughout the training and validation phases of a YOLO model over the course of 100 epochs. The graphs located in the top left and the centre illustrate the training losses associated with bounding box regression (train/box_loss) and objectiveness scores (train/obj_loss), both of which exhibit a consistent decline as the epochs advance. This trend signifies that the model is enhancing its capability to accurately predict bounding boxes and assess object confidence. Conversely, the graph labelled train/cls_loss remains consistently at zero, indicating that class label information is not being utilized during this training phase, likely due to the model being focused on object detection without classification. In the validation losses displayed in the bottom row, both val/box_loss and val/obj_loss demonstrate a downward trajectory, suggesting that the model generalizes effectively to previously unseen data. Similar to the training loss,

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

val/cls_loss also remains at zero. The precision and recall metrics reach a state of stability following an initial increase, which implies that the model is becoming more reliable in detecting objects and reducing false positives as training continues. The metrics/mAP_0.5 graph shows a swift rise followed by a plateau, indicating the model's proficiency in accurately detecting objects at an IoU threshold of 0.5. Additionally, metrics/mAP_0.5:0.95, which assesses performance across a spectrum of IoU thresholds, also shows an upward trend and subsequent stabilization, reflecting strong performance across varying degrees of overlap. In summary, the observed trends indicate a successful convergence of the model, characterized by enhanced detection capabilities and minimal overfitting, as evidenced by the alignment of training and validation losses and the stabilization of metrics. The lack of classification loss further reinforces that this task is exclusively centered on object detection.

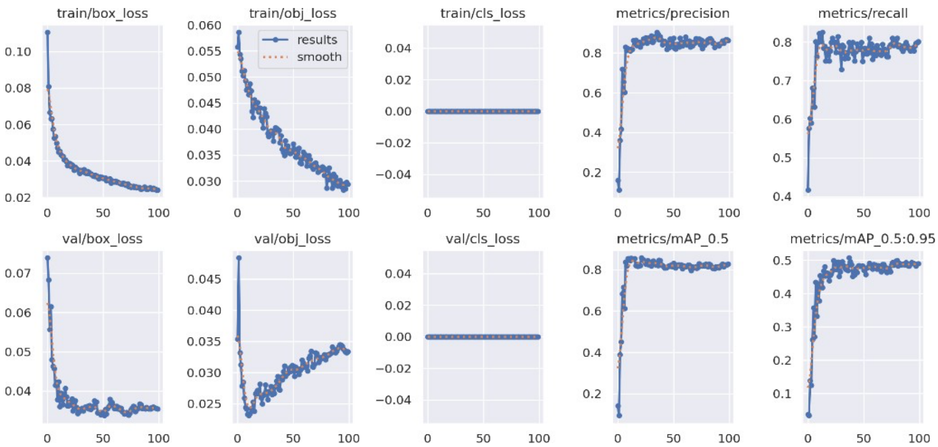

Figure 6. Performance Result YOLOv5

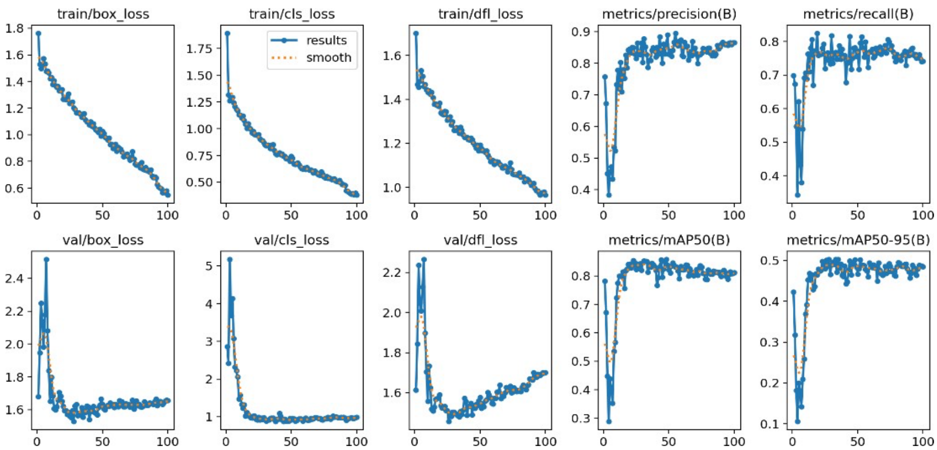

Figure 7. Performance Result YOLOv8

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 7 is the performance result of YOLOv8, the illustration presents the training and validation metrics of a YOLO model across 100 epochs. The upper section details the training metrics, which encompass box loss, classification loss, and distribution focal loss (DFL). A consistent decline in these losses signifies the model's enhancement in object localization, accurate classification, and refinement of bounding box predictions as the training advances. The precision and recall metrics in this section exhibit initial fluctuations during the early epochs but eventually stabilize, indicating improved detection accuracy and a reduction in false negatives. The lower section is dedicated to validation metrics. The box loss, classification loss, and DFL for the validation set demonstrate comparable decreasing trends to those observed in the training metrics, implying effective generalization without indications of overfitting. The Mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 (mAP50) and across a spectrum of IoU thresholds (mAP50-95) shows consistent improvement throughout the training process, ultimately stabilizing at elevated values. This reinforces the model's capability to maintain accurate detection and localization across various IoU thresholds. In summary, the plots reflect successful learning, convergence, and dependable performance on previously unseen data.

Figure 8. Human Detection Result YOLOv5

Figure 8 shows the results of human detection using the YOLOv5 model, trained in Google Collab. The model is configured to detect a single class, "person." Each

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

detected person in the image is highlighted with bounding boxes labeled as "person," showcasing the detection output. The detection demonstrates the model's ability to identify people in a variety of settings and clothing styles.

Figure 9 . Human Detection Result YOLOv8

Figure 9 depicts the results of human detection performed using the YOLOv5 model. The model was trained in Google Collab and configured to detect only one class, "person." In this visualization, detected individuals are highlighted with bounding boxes labelled as "person," using a blue colour scheme. The results demonstrate the model's capacity to accurately identify people in diverse scenarios and environments. From the two batches above (Figure 8 and Figure9), it is obtained that YOLOv8 is better than YOLOv5 because according to the data above, YOLOv5 is only superior in recall value while YOLOv8 is superior in precision and F1-score value.

3.3. Confusion Matrix

These matrices revealed minimal off-diagonal elements, signifying a low occurrence of misclassification events [25]. The dataset used in this research comes from Roboflow, specifically an object detection dataset for human detection. This research uses Google Collab for training and validating images, with a total of 100 epochs. This is a confusion matrix resulting from training in Google Collab with a total of 100 epochs. We conduct training on a dataset for 100 epochs using Google Collab to provide the model with sufficient iterations for effective data learning, to achieve a balance between underfitting and overfitting, and to optimize its performance within the available resources of the platform.

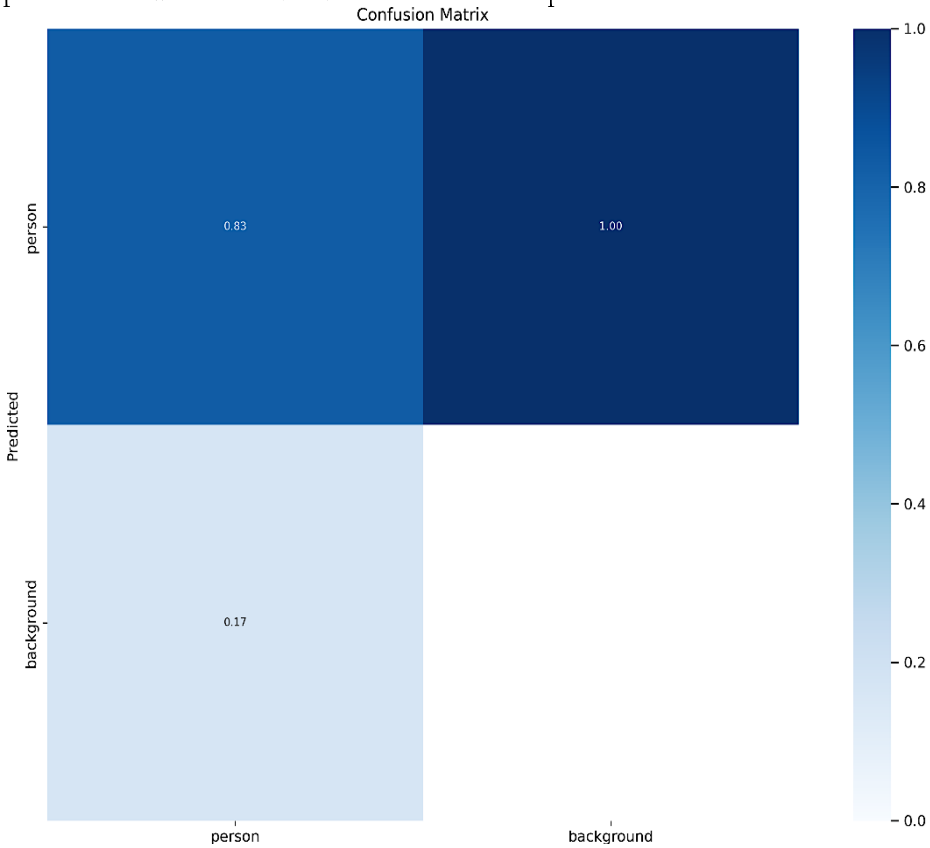

Figure 10. Confusion Matrix YOLOv5

Figure 10 shows the confusion matrix of YOLOv5 at the final epoch, with values of TP = 0.83, FP = 1, and FN = 0.17. These values result from detecting humans in images during the validation phase. The precision value of equation 1 is $\frac{0.83}{0.83+1}$ = 0.4535 or 45.35%. The recall value of equation 2 is $\frac{0.83}{0.83+0.17}$ = 0.83 or 83%. The F1-score value of equation 3 is $\frac{0.7528}{0.4535+0.83}$ = 0.5865 or 58%. Figure 11 shows confusion matrix of YOLOv8, the result shows that TP = 223, FP=71, and FN = 46. These values result from detecting humans in images during the validation stage. The precision value of equation 1 is $\frac{223}{223+71}$ = 0.7585 or 75.85%. The recall value of equation 2 is $\frac{223}{223+46}$ = 0.8289 or 82.89%. The F1-score value of equation 3 is $\frac{1.2574}{1.5874}$ = 0.7921 or 79.21%.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

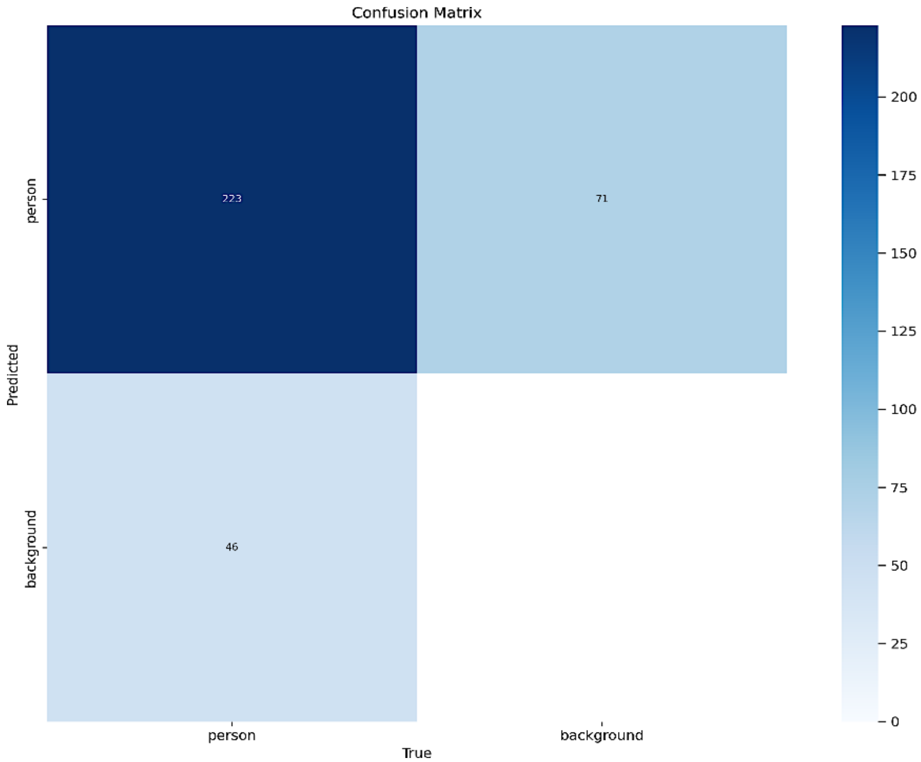

Figure 11. Confusion Matrix YOLOv8

The comparation of performance values of the YOLOv5 and YOLOv8 models are summarized in Table 1. The YOLOv8 model excels in value precision and F1- Score and the YOLOv5 model excels in value recall.

Table 1. YOLOv5 and YOLOv8 Model Performance

Architecture	Precision	Recall	F1-Score
YOLOv5	0.4535	0.83	0.5865
YOLOv8	0.7585	0.8289	0.7921

4. CONCLUSION

Based on the result of the experiment, shows that YOLOv8 has better performance based on data Precision 0.7585 or 75.85% dan F1-Score 0.7921 or 79.21% otherwise in Recall YOLOv8 almost same with YOLOv5. The findings indicate that although both models are competent, but YOLOv8 has better performance than YOLOv5 regarding Precision, F1-score and Recall. From these result showcasing that YOLOv8 enhanced accuracy and efficiency in human detection tasks. YOLOv5, on the other hand, has a minor advantage in recall,

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

making it potentially preferable for certain situations where identifying all possible instances is crucial, though this comes with a decrease in precision. The contribution of this study can emphasize the progress in AI and computer vision, concentrating on human or object recognition or detection.

REFERENCES

[1] D. P. Sari and A. H. Mirza, "The detection of face recognition as employee attendance presence using the YOLO algorithm (you only look once)," *Jurnal Darma Agung*, vol. 30, no. 3, pp. 41-50, 2022.
[2] H. Haeruddin, H. Herman, and P. P. Hendri, "Pengembangan aplikasi emoticon recognition dan facial recognition menggunakan algoritma local binary pattern histogram (LBPH) dan convolutional neural network (CNN)," *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 49–55, 2023, doi: 10.54914/jtt.v9i1.613.
[3] E. Utama, F. Yapputra, and G. Gasim, "Identifikasi jenis mangga berdasarkan bentuk menggunakan fitur HOG dan jaringan syaraf tiruan," *Jurnal Ilmiah Informatika Global*, vol. 9, no. 1, 2018.
[4] S. Marappan, P. Kuppuswamy, R. John, and N. Shanmugavadivu, "Human detection in still images using HOG with SVM: A novel approach," in *Proc. Second Int. Conf. on Innovative Computing and Cutting-edge Technologies (ICICCT 2020)*, Springer International Publishing, 2021, pp. 385-397, doi: 10.1007/978-3-030-65407-8_33.
[5] T. A. Nguyen, T. Q. Tran-Thi, D. H. Bui, and X. T. Tran, "FPGA-based human detection system using HOG-SVM algorithm," in *2023 Int. Conf. on Advanced Technologies for Communications (ATC)*, IEEE, 2023, pp. 72-77.
[6] Y. He, "A novel low-resource consumption and high-speed hardware implementation of HOG feature extraction on FPGA for human detection," *Integration*, vol. 97, 2024, doi: 10.1016/j.vlsi.2024.102208.
[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2016, doi: 10.1109/TPAMI.2015.2437384.
[8] Y. Wang, J. Wu, and H. Li, "Human detection based on improved Mask R-CNN," in *J. Phys.: Conf. Ser.*, Institute of Physics Publishing, vol. 1575, no. 1, Jul. 2020, doi: 10.1088/1742-6596/1575/1/012067.
[9] K. R. Akshatha, A. K. Karunakar, S. B. Shenoy, A. K. Pai, N. H. Nagaraj, and S. S. Rohatgi, "Human detection in aerial thermal images using faster R-CNN and SSD algorithms," *Electronics (Switzerland)*, vol. 11, no. 7, Apr. 2022, doi: 10.3390/electronics11071151.
[10] S. S. Sumit, J. Watada, A. Roy, and D. R. A. Rambli, "In object detection deep learning methods, YOLO shows supremum to Mask R-CNN," in *J. Phys.: Conf. Ser.*, Institute of Physics Publishing, vol. 1529, no. 4, Jun. 2020, doi: 10.1088/1742-6596/1529/4/042086.

[11] K. Boudjit, "Human detection based on deep learning YOLO-v2 for real-time UAV applications," *J. Exp. Theor. Artif. Intell.*, vol. 34, no. 3, pp. 527–544, 2022, doi: 10.1080/0952813X.2021.1907793.
[12] M. D. Klein, Z. J. Edel, C. D. Packard, J. N. Hendrickson, A. C. Levanen, and P. L. Rynes, "Automatic detection and recognition of humans with YOLO-based deep learning using MuSES-generated EO/IR synthetic imagery," in *Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications II*, SPIE, vol. 13035, pp. 66-82, Jun. 2024, doi: 10.1117/12.3013911.
[13] A. Raza, M. H. Yousaf, and S. A. Velastin, "Human fall detection using YOLO: a real-time and AI-on-the-edge perspective," in *2022 12th Int. Conf. on Pattern Recognition Systems (ICPRS)*, IEEE, 2022, pp. 1-6, doi: 10.1109/ICPRS54038.2022.9854070.
[14] P. A. Cahyani, M. Mardiana, P. B. Wintoro, and M. A. Muhammad, "Sistem perhitungan kendaraan menggunakan algoritma YOLOv5 dan DeepSORT," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, May 2024, doi: 10.28932/jutisi.v10i1.7519.
[15] D. I. Mulyana and M. A. Rofik, "Implementasi deteksi real time klasifikasi jenis kendaraan di Indonesia menggunakan metode YOLOV5," *Jurnal Pendidikan Tambusai*, vol. 6, no. 3, 2022, pp. 13971-13982.
[16] N. P. Motwani and S. S, "Human activities detection using deep learning technique- YOLOv8," *ITM Web of Conferences*, vol. 56, p. 03003, 2023, doi: 10.1051/itmconf/20235603003.
[17] N. Hidayat, S. Wahyudi, and A. A. Diaz, "Pengenalan individu melalui identifikasi wajah menggunakan metode You Only Look Once (Yolov5)," *UNEJ e-Proceeding*, pp. 85-98, 2022.
[18] N. Ma'muriyah, A. Yulianto, and Lili, "Design prototype of audio guidance system for blind by using raspberry pi and fuzzy logic controller," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012024.
[19] A. Setiyadi, E. Utami, and D. Ariatmanto, "Analisa kemampuan algoritma YOLOv8 dalam deteksi objek manusia dengan metode modifikasi arsitektur," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 7, no. 2, pp. 891-901, 2023.
[20] P. D. Manalu and Z. Situmorang, "Face recognition using support vector machine (SVM) and backpropagation neural network (BNN) methods to identify gender on student identity cards," in *Proc. of International Conference on Information Science and Technology Innovation (ICoSTEC)*, vol. 3, no. 1, Feb. 2024.
[21] Z. Ahmad, S. Yaacob, R. Ibrahim, and W. F. Wan Fakhruddin, "The review for visual analytics methodology," in *HORA 2022 - 4th International Congress on Human-Computer Interaction, Optimization and Robotic Applications*, 2022, doi: 10.1109/HORA55278.2022.9800100.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[22] Y. A. Suwitono and F. J. Kaunang, "Implementasi algoritma convolutional neural network (CNN) untuk klasifikasi daun dengan metode data mining SEMMA menggunakan Keras," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 6, no. 2, pp. 109–121, 2022, doi: 10.31603/komtika.v6i2.8054.

[23] P. D. Manalu and Z. Situmorang, "Face recognition using support vector machine (SVM) and backpropagation neural network (BNN) methods to identify gender on student identity cards," in *Proc. of International Conference on Information Science and Technology Innovation (ICoSTEC)*, vol. 3, no. 1, Feb. 2024.

[24] D. Arifadilah and A. Pambudi, "Sunda script detection using You Only Look Once algorithm," *J. Artif. Intell. Eng. Appl. (JAIEA)*, vol. 3, no. 2, pp. 606-613, 2024.

[25] P. K. Pativada, "Real-time detection and classification of plant seeds using YOLOv8 object detection model," 2024.