Journal of Applied Research and Technology

www.jart.icat.unam.mx

ICAT Instituto de Ciencias Aplicadas y Tecnología

Journal of Applied Research and Technology 24 (2026) 376-387

Original

Face Detection and Recognition in CCTV Security: A Comparative Study of YOLOv5 and YOLOv8

Muhammad Rif'an[1] • Suci Dwijayanti[2]*
Bhakti Yudho Suprapto[2] • Muhammad Yulwi Alwan[2]

[1]*Information System, Universitas Terbuka, Jakarta, Indonesia*
[2]*Department of Electrical Engineering, Universitas Sriwijaya, Inderalaya, Indonesia*

Received: 27 06 2024; Accepted: 27 04 2025
Available: 30 04 2026

Abstract: Security systems traditionally rely on CCTV for monitoring spaces accessible only to authorized personnel, yet they struggle with face detection and recognition at distances beyond a few meters. This limitation hampers their effectiveness in enhancing room security. This study addresses this challenge by developing a remote facial recognition system utilizing CCTV cameras to identify faces from 1-3 meters away. We employed YOLOv5 and YOLOv8 algorithms, testing pre-trained models of varying sizes (M and X) to improve detection accuracy. The training phase involved 200 epochs with a batch size of 32, yielding mean Average Precision (mAP) scores of 82.7%, 83%, 85%, and 85.2% for YOLOv5m, YOLOv5x, YOLOv8m, and YOLOv8x, respectively. Offline evaluations demonstrated average accuracy rates of 94%, 95%, 90%, and 91%. Online testing, conducted under varying conditions with 1-3 faces visible, showed YOLOv5x achieving an accuracy of 87.8%, compared to 80.9% for YOLOv8x. The results indicate that while single-face recognition is quick and accurate, performance declines with multiple faces in view. This research offers a promising solution to enhance room security through effective facial recognition at a distance, highlighting the potential of improved surveillance technology in secure environments.

Keywords: **CCTV, YOLOv5, YOLOv8, comparative study, accuracy**

*Corresponding author.
E-mail address: sucidwijayanti@ft.unsri.ac.id (Suci Dwijayanti).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

Security systems are crucial components in room surveillance. These systems typically employ various approaches, such as personal identification numbers (PINs), cards, or biometrics. Biometrics offer distinct advantages over other methods. Various biometrics, including fingerprints, voice recognition, and facial recognition, are utilized. Facial recognition is the most widely used biometric modality (Fahad et al., 2017) and can be easily detected using closed-circuit television cameras (CCTV) for location surveillance to identify individuals (Kanyal et al., 2020). CCTV cameras can efficiently facilitate the surveillance process (Halawa et al., 2019).

The face captured by CCTV needs to be detected and recognized. Various studies have employed different approaches to identify and recognize faces in CCTV surveillance. For instance, Nurhopipah & Harjoko (2018) proposed a motion detection and face recognition system using accumulated images for motion detection and Speeded-Up Robust Features (SURF) and Principal Component Analysis (PCA) for face feature extraction. The study utilized a Haar cascade classifier to recognize faces.

In another approach, Aung et al. employed a VGG16 pre-trained convolutional neural network (CNN) for feature extraction and You Only Look Once (YOLO) for object detection based on these features. However, this method was not implemented in real time using the Face Detection Dataset and Benchmark (FDDB).

Additionally, Ullah et al. (2022) used CNNs for automatic face recognition in CCTV systems, comparing them with various machine learning methods, including k-nearest neighbors (KNN), decision trees, and random forests. Although this study used only single-frame images, it highlighted the effectiveness of CNNs for facial recognition tasks.

Furthermore, Sino & Areni (2019)face recognition with low-resolution data from CCTV is needed. In this study, Viola Jones, Gabor filter and Support Vector Machine (SVM combined Viola-Jones, Gabor filters, and support vector machines to detect faces, extract features, and classify them for CCTV usage. Another approach used Faster R-CNN with the Inception-V2 architecture for face recognition in CCTV cameras, as proposed by Halawa et al. (2019). Son et al. (2020) implemented the CCTV system as an attendance tool and tested it with various classifiers.

Among the techniques mentioned earlier, YOLO models have received significant attention due to their ability to perform real-time object detection. Several studies have employed YOLO for object detection, demonstrating its versatility. For instance, Mun & Lee (2022) utilized Tiny YOLOv3 to verify visitors using CCTV cameras. Abhinand et al (2021) detected moving objects by leveraging YOLOv3's capabilities, while Menaka & Yogameena (2021) implemented the YOLOv2 framework to detect blurred images improved by discrete wavelet transform. Wang et al. (2023) used YOLOv4 on CCTV footage to detect small objects such as weapons. YOLOv5 has been employed to detect mask wearing (Guo et al., 2022)and proper wearing of a mask can hinder the spread of the virus. However, complex factors in natural scenes, including occlusion, dense, and small-scale targets, frequently lead to target misdetection and missed detection. To address these issues, this paper proposes a YOLOv5-based mask-wearing detection algorithm, YOLOv5-CBD. Firstly, the Coordinate Attention mechanism is introduced into the feature fusion process to stress critical features and decrease the impact of redundant features after feature fusion. Then, the original feature pyramid network module in the feature fusion module was replaced with a weighted bidirectional feature pyramid network to achieve efficient bidirectional cross-scale connectivity and weighted feature fusion. Finally, we combined Distance Intersection over Union with Non-Maximum Suppression to improve the missed detection of overlapping targets. Experiments show that the average detection accuracy of the YOLOv5-CBD model is 96.7%—an improvement of 2.1% compared to the baseline model (YOLOv5, to detect faces in security surveillance scenes (Xu et al., 2021), and, specifically, as a face detector (Qi et al., 2022). Furthermore, Majeed et al. (2022) investigated the effectiveness of YOLOv5 in real-time face recognition using the FDDB dataset. Moreover, Sholahuddin et al. (2023)especially CCTV systems, requires fast and accurate face detection. Object detection models with slow inference times are ineffective in real-time. This study addresses this challenge by improving the inference speed of the YOLOv8 model, a leading object detection framework known for its accuracy and speed. We focus on pruning the model's architecture, particularly the P5 head section, which detects larger objects. According to Bochkovskiy's 2020 research, this modification enhances the model's performance specifically for medium and small objects in CCTV footage. The standard YOLOv8 model and its modified version were compared for inference time, mean Average Precision (mAP used YOLOv8 for real-time object detection.

Based on these methods, YOLO models have garnered significant attention for their ability to perform real-time object detection. Specifically, YOLOv5 and YOLOv8 represent advancements in the YOLO architecture, each

offering unique features and improvements over previous iterations. Comparative studies of YOLO models have been conducted (Khalfaoui et al., 2022), although they utilized a dataset (Penn-Fudan) rather than real-time data. More recent studies have compared YOLOv5 and YOLOv7 for object detection using the Google Open Images Dataset (Olorunshola et al., 2023) and compared Viola-Jones and YOLOv3 for real-time face detection using the COCO_MS database (Obaida et al., 2022). Notably, previous comparative studies have not been conducted in real-time.

Currently, many studies focus on face detection and recognition at close range, whereas CCTV cameras are often used to capture images from a distance. Moreover, existing methods have not been directly applied to security systems. Therefore, this paper presents a comparative study of YOLOv5 and YOLOv8 for face detection and recognition within a CCTV-based security system. Face detection and recognition are crucial tasks in surveillance applications, where accuracy and efficiency are essential. This study aims to evaluate the performance, robustness, and practical suitability of these models under controlled conditions and real-world scenarios, providing valuable insights for security implementations. The contributions of this paper can be summarized as follows:

- A comparative analysis of YOLOv5 and YOLOv8 is presented for real-time face detection and recognition in CCTV-based applications.
- This study uses images of real objects as primary data.
- Testing is conducted using real-world scenarios, involving multiple objects in a single frame.

The remainder of this paper is organized as follows: Section 2 provides an overview of the methodology and experimental setup used to evaluate YOLOv5 and YOLOv8, detailing the specific procedures employed in this study. Section 3 presents the experimental results and analyzes their implications, followed by a discussion section that highlights the key findings. Finally, Section 4 concludes the paper by summarizing the main outcomes and suggesting potential avenues for future research.

2. Method

The experiment in this study was conducted as depicted in Figure 1. A Reolink RLC-410W CCTV camera was installed, as shown in Figure 2, to capture data under two distinct lighting conditions: bright and less bright. The dataset was collected within a specific angle range and distance to ensure effective data collection. To ensure

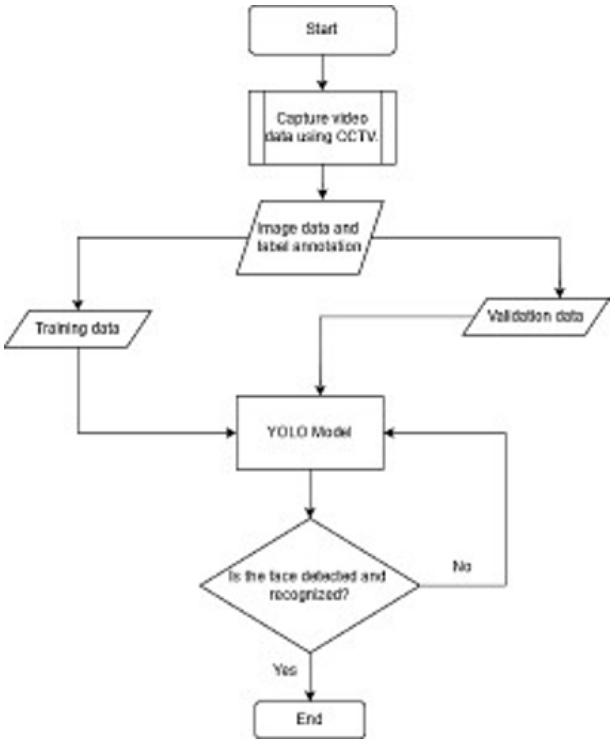


Figure 1. Flowchart of face detection and recognition system.

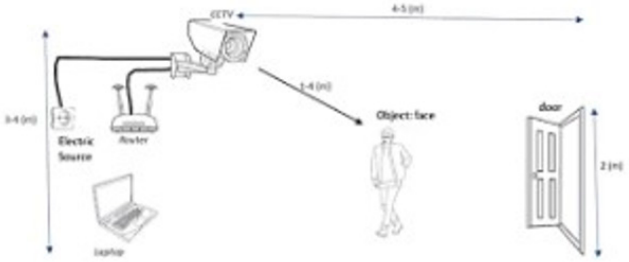


Figure 2. Sketch of installed CCTV for dataset collection.

accurate labeling, faces were required to be uncovered (i.e., without glasses or hats). The video data was then converted into images, which were annotated to label each class. The datasets were divided into training (80%) and validation sets (20%). The training data were used to train the YOLOv8 and YOLOv5 models, while the models obtained from training were tested using the validation data. The success of the models was measured by their ability to detect and recognize faces.

The primary data used in this study was collected using CCTV cameras from 66 participants. Each face was recorded for 15-25 seconds, with a distance of 1-3 meters between the object and the CCTV camera. The recorded

video was then converted into images with a resolution of 2560 x 1440 pixels. Each class of images consisted of 70 images, for a total of 4,620 images processed in this study.

In this study, the YOLO algorithm was chosen for face detection and recognition due to its effectiveness. Two versions of YOLO were used: YOLOv5 and YOLOv8. The architectures of these models are depicted in Figures 3 and 4, respectively. YOLOv5 uses a modified CSPDarknet53 backbone to extract features from the input image. It comprises a stem layer, convolutional layers, and Spatial Pyramid Pooling (SPPF) layers for multi-scale processing. The network also incorporates a modified Cross-Stage Partial Path Aggregation Network (CSP-PAN) module in the neck section, similar to YOLOv3's head architecture (Terven et al., 2023).

YOLOv8 leverages a backbone similar to YOLOv5, with modifications to the CSPLayer.

Now referred to as the C2f module. The C2f module combines high-level features with contextual information to enhance detection accuracy by incorporating partial bottleneck structures with two convolutions. In contrast to YOLOv5, YOLOv8 employs an anchor-free architecture with separate heads for object detection, classification, and regression. This design allows each branch to focus on its specific task, ultimately improving the model's overall accuracy. In the output layer of YOLOv8, sigmoid and softmax activation functions are used for the objective score and class probabilities, respectively. The objective score represents the probability that a bounding box contains an object, while class probabilities represent the likelihood of the object belonging to each possible class.

The evaluation metric is accuracy, which is calculated as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (1)$$

where True Positive (TP) is the number of data points that are actually positive and correctly predicted as positive by the system. False Positive (FP) is the number of data points that are actually negative but incorrectly predicted as positive by the system. True Negative (TN) is the number of data points that are actually negative and correctly predicted as negative by the system. False Negative (FN) is the number of data points that are actually positive but incorrectly predicted as negative by the system.

Meanwhile, the training metrics consist of tran/box_loss, train/obj_loss, train/cls_loss, precision, recall, val/box_loss, val/obj_loss, val/cls_loss, mean average precision (mAP), and mAP50-95.

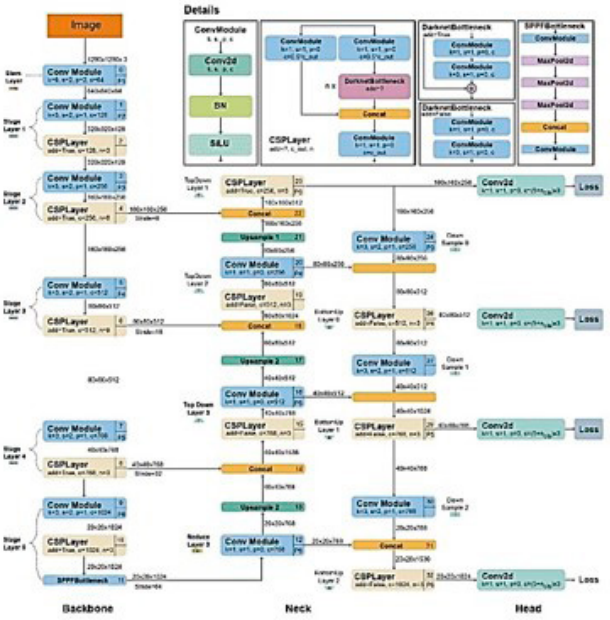


Figure 3. Architecture of YOLOv5 (Terven et al., 2023).

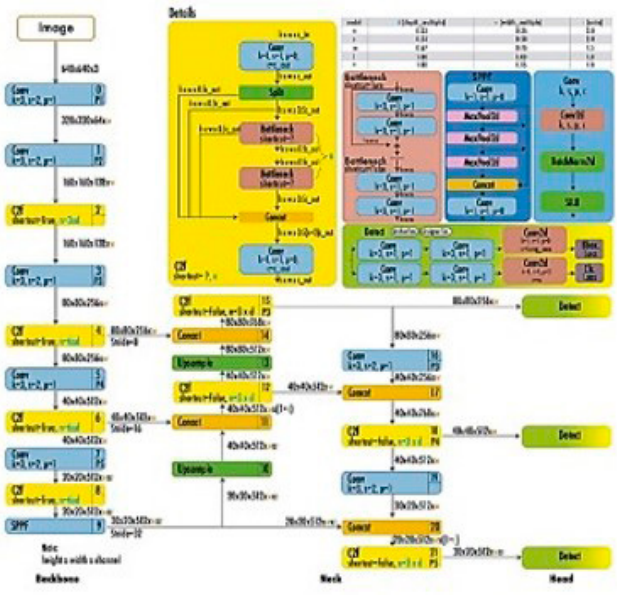


Figure 4. Architecture of YOLOv8 (Terven et al., 2023).

3. Results and Discussion

The total dataset used in this study consists of 4,620 face images. Prior to training, each image was preprocessed by resizing it from 2560x1440 pixels to 640x640 pixels.

Annotation was performed using Roboflow al., 2023), in which each face in the image was labeled with a bounding box. An example of the annotation process is shown in Figure 5. The resulting annotations provided information on class, x-coordinate, y-coordinate, width, and height, as illustrated in Figure 5.

After annotation labeling is performed on the image, it produces label files in `.txt` format that contain object class values and bounding box coordinates (x, y, width, height). For example, the value `0 0.45 0.55 0.25 0.35` represents Class 0, with bounding box coordinates (x, y) at `(0.45, 0.55)` and a width and height of `(0.25, 0.35)`.

The dataset for each class, consisting of 70 samples, is split into training and validation data with an 80/20 ratio. During training, a Yet Another Markup Language (YAML) file is required to specify the paths to the training and validation data directories, the number of classes to be trained, and the YOLO model size. The class values in the labels are sorted alphabetically.

There are two types of YOLO models employed in this study: YOLOv5 and YOLOv8. Each model variant, denoted by 'm' and 'x', differs in its architecture. Specifically, version 'm' has fewer layers than version 'x', resulting in a faster inference speed and lower computation requirements. The detailed specifications of each YOLOv5 and YOLOv8 model used in this study can be found in Table 1

Table 1. The detailed parameters of YOLOv5 and YOLOv8.

Type of YOLO	Num of layers	Num of params	Num of gradients	Giga floating point operations per second (GFLOPs)
YOLOv5m	291	21133983	21133983	49.1
YOLOv5x	445	86655199	86655199	206.0
YOLOv8m	295	25894534	25894518	-
YOLOv8x	365	68216166	68216150	-

During training, the hyperparameters used for each architecture were identical, including a batch size of 32 and 200 epochs. The results of the trained models for YOLOv5 and YOLOv8 can be seen in Figures 6 to 9.

In Figures 6 and 7, it can be observed that the losses and metrics obtained from the training are satisfactory. The decreasing values of box_loss, obj_loss, and cls_loss demonstrate the effectiveness of the training process. Throughout both the training and validation phases, these loss values consistently decrease from epoch 0 to 200. Concurrently, precision, recall, mAP_0.5, and mAP_0.5:0.95 metrics steadily increase, indicating that

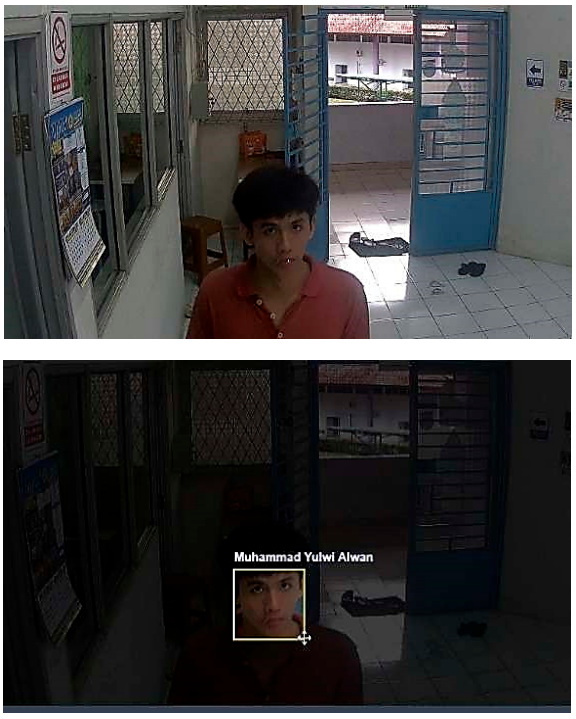


Figure 5. The original data (a) and the annotated and labeled data (b).

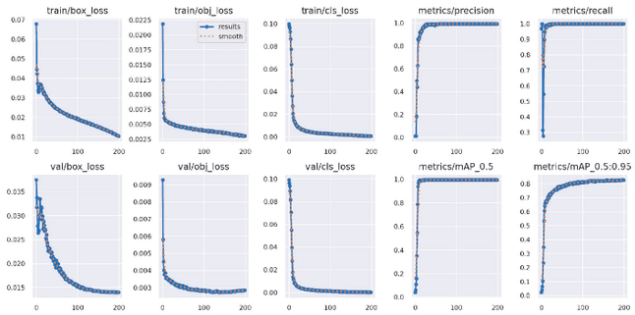


Figure 6. Training metrics and loss of YOLOv5m.

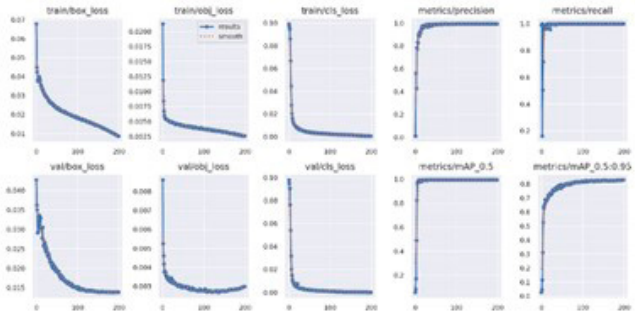


Figure 7. Training metrics and loss of YOLOv5x.

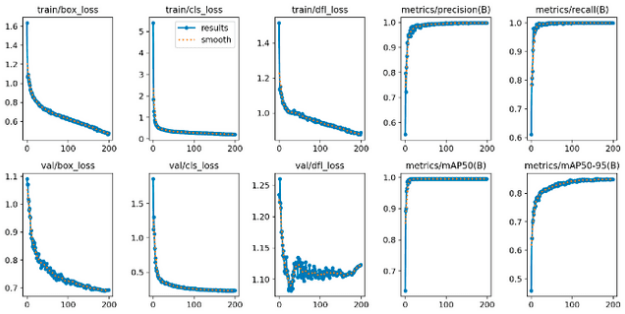


Figure 8. Training metrics and loss of YOLOv8m.

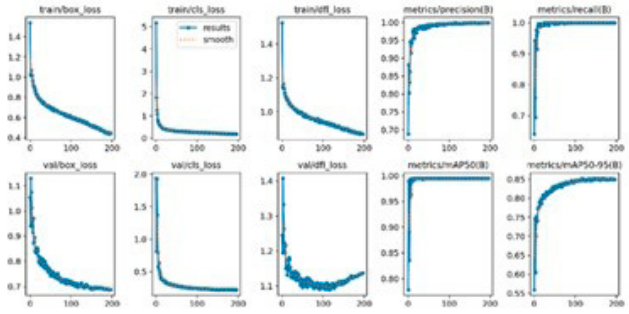


Figure 9. Training metrics and loss of YOLOv8x.

the model performs well. Notably, YOLOv5x (as shown in Figure 7) exhibits similar trends in losses and metrics as YOLOv8 (Figure 6), with consistently decreasing loss values and increasing metric values throughout both phases.

Figure 8 presents the results of training YOLOv8m, where the decreasing trends in box_loss and cls_loss throughout both the training and validation phases indicate good performance. In contrast, the dfl_loss section during the training phase shows consistently decreasing values, whereas the validation phase shows unstable, higher values than during training. This suggests that overfitting has occurred, affecting the suboptimal results achieved by the model. On the other hand, all metric values show excellent performance, with consistent increases from epoch 0 to 200.

Figure 9 displays similar trends in losses (box_loss, cls_loss) for YOLOv8x as seen in YOLOv5m, with both training and validation phases exhibiting consistently decreasing values. However, dfl_loss shows distinct patterns during training: it initially decreases, then increases, surpassing its training-phase values. This indicates overfitting, which negatively impacts the model's performance. Despite this, the metric values obtained for YOLOv8x remain good, as they consistently increase over the 200 epochs.

As depicted in Figure 10, we observe the accuracy results for four trained models: YOLOv5m, YOLOv5x, YOLOv8m, and YOLOv8x. Notably, all four models exhibit a steady increase in accuracy from epoch 0 to 200, indicating good performance. Initially, YOLOv8 models demonstrate faster convergence toward high accuracy, reaching values close to 50% in the early epochs. This suggests that YOLOv8 has better computational performance for achieving high accuracy.

However, as training progresses, the four models converge and show similar accuracy, with little distinction between them until the end of the graph. The final accuracy values achieved by these models after 200 epochs of training are shown in Table 2.

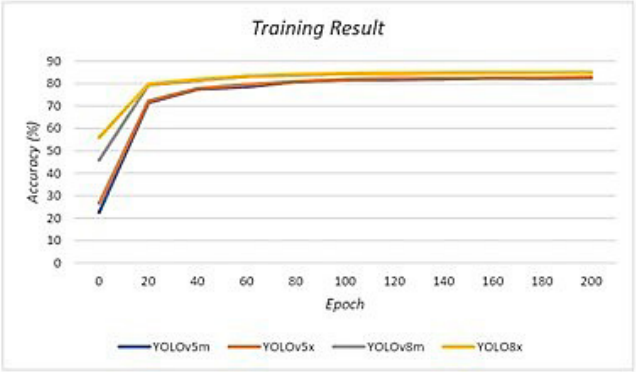


Figure 10. Training accuracy of four models.

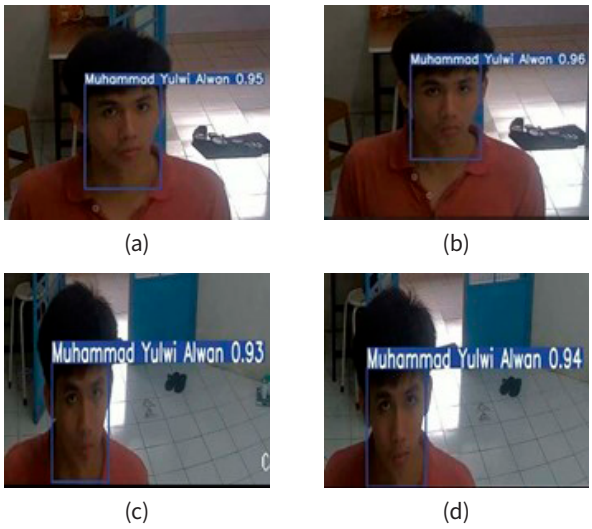
Table 2. The best training accuracy for each model.

	Training accuracy (mAP_0.5:0.95)			
Epoch	YOLOv5m	YOLOv5x	YOLOv8m	YOLOv8x
0-200	0.827	0.83	0.85	0.852

From Table 2, we observe that both YOLOv8 sizes, X and M, achieve higher accuracy than YOLOv5. Notably, the accuracy difference between the M and X architectures is not significant for face detection and recognition tasks. This suggests that the X-sized model, despite having the highest number of layers and parameters, requires longer computation times due to its design focus on high accuracy at the expense of slower inference speed and increased computational requirements. In contrast, the M-sized model has fewer layers and parameters, yet it achieves faster inference and lower computational cost. The M size strikes a balance between accuracy and speed, making it an attractive choice for applications that require a good trade-off between the two.

3.1 Offline Testing Data

A sample of offline testing data is shown in Figure 11 for each of the YOLOv5 and YOLOv8 models.



(a) (b) (c) (d)

Figure 11. The results of offline testing of the classes: (a) YOLOv5m; (b) YOLOv5x; (c) YOLOv8m; (d) YOLOv8x.

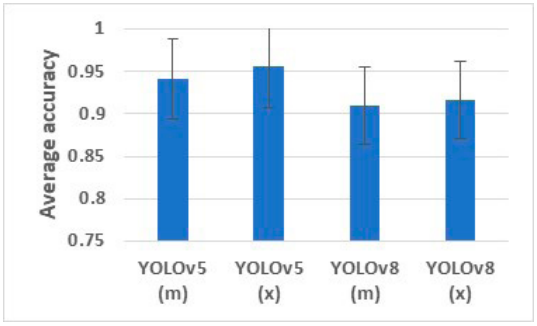


Figure 12. Average accuracy of different types of YOLOv5 and YOLOv8.

Table 3 presents the accuracy results in terms of the probability of confidence values obtained for all models across 66 classes. Notably, each class exhibits varying accuracy across model sizes, with the X-sized model standing out as the most effective at achieving high accuracy for both YOLOv5 and YOLOv8. The confidence interval of YOLOv5 is statistically significant compared to YOLOv8, as shown in Table 3 and Figure 12. This is further supported by the p-value between two different types of YOLOv5 models, indicating a statistically significant difference between the compared groups. While the X-sized model demonstrates better accuracy than other versions, it requires substantial computational resources to run. These offline testing results are consistent with the findings presented in Table 2, which reports the mAP0.50:0.95 values (i.e., accuracy) for the trained models.

Based on Table 3, for the class "Muhammad Yulwi Alwan" in the YOLOv5 model, the M-sized model achieves the highest accuracy of 0.95, while the X-sized model achieves a slightly higher value of 0.96. Similarly, for YOLOv8, the M-sized model achieves an accuracy of 0.93, and the X-sized model achieves 0.94. The X-sized model has a higher average testing accuracy than the M-sized model, but the M-sized model could achieve similar or better accuracy in this offline testing. For example, in test 21 for YOLOv5, both the M- and X-sized models achieve the same accuracy of 0.97, and in test 7 for YOLOv8, the M-sized model achieves a higher accuracy of 0.95, while the X-sized model achieves a lower accuracy of 0.92. Overall, the average confidence level accuracy for 66 classes in YOLOv8 is slightly lower compared to YOLOv5, at 90% for the M-sized model and 91% for the X-sized model. This suggests that YOLOv5 outperforms YOLOv8, but the difference in accuracy is not statistically significant given that YOLOv8 builds upon the YOLOv5 framework. The lower accuracy of YOLOv8 may also be due to YOLOv5's more established architecture, which performs well across various datasets, including faces. In addition, YOLOv5 is better suited to scenarios with limited resources or time constraints (Casas et al., 2024). YOLOv5 may also achieve better accuracy because it has fewer parameters, which can affect its speed and face recognition accuracy.

3.2 Online Testing Using CCTV

Online testing was conducted in three locations within the laboratory: the left, middle, and right sides, with CCTV cameras positioned to face the door. This testing aimed to evaluate whether the CCTV placement positions were

Table 3. The average accuracy for 66 classes in offline testing.

	YOLOv5		YOLOv8	
	Medium (m)	Extra large (x)	Medium (m)	Extra large (x)
Average acc.	94%	95%	90%	91%
Confidence interval	0,941515152 ± 0,00448635	0,955454545 ± 0,003508034	0,909393939 ± 0,00913606	0,915757576 ± 0,005282357
p-value	2,94398E-06		0,230669315	

Table 4. The results of online testing under conditions of light on and off with a single face object to be detected.

No	Condition	Sample	YOLOv5x	YOLOv8x
1	Light on – object in the left position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Misrecognized
		Sample 3	Recognized	Recognized
2	Light off – object in the left position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Misrecognized	Recognized
3	Light on – object in the middle position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Unrecognized
		Sample 3	Unrecognized	Recognized
4	Light off – object in the middle position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Recognized	Recognized
5	Light on – object in the right position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Misrecognized	Recognized
6	Light off – object in the right position	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Unknown	Undetected

Table 5. The results of online testing under conditions of light on and off with two face objects to be detected.

No	Condition	Sample	YOLOv5x	YOLOv8x
1	Light on – 2 objects in the left lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Undetected
2	Light on – 2 objects in the left in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
3	Light off – 2 objects in the left lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Detected
4	Light off – 2 objects in the left in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Undetected
5	Light on – 2 objects in the center lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Misrecognized
6	Light on – 2 objects in the center in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
7	Light off – 2 objects in the center lined up	Sample 1	Recognized	Undetected
		Sample 2	Recognized	Recognized
8	Light off – 2 objects in the center in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
9	Light on – 2 objects in the right lined up	Sample 1	Misrecognized	Recognized
		Sample 2	Undetected	Undetected
10	Light on – 2 objects in the right in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
11	Light off – 2 objects in the right lined up	Sample 1	Undetected	Undetected
		Sample 2	Recognized	Recognized
12	Light off – 2 objects in the center in rows	Sample 1	Undetected	Undetected
		Sample 2	Undetected	Undetected

optimal. The online test also considered various conditions, including normal lighting with the lights on and the laboratory door open, as well as dark conditions with the lights off and the laboratory door closed. Additionally, the test examined scenarios with varying numbers of people in the frame: 1 person, 2 people, and 3 people positioned in a line or rows. Furthermore, an Unknown class condition was included for facial objects not present in the training data and with confidence probabilities below 0.70. This unknown-individual detection was designed to assess the robustness of the face detection and recognition system built as a room security system in detecting individuals not recognized by the system.

In the online testing results for face detection and recognition, satisfactory performance was observed under various conditions, as shown in Table 4. Both models were able to recognize single faces from distances of 1-2 meters from the CCTV camera position. However, light intensity emerged as a critical factor in face detection and recognition, particularly when the face was not centered on the CCTV, such as in the condition where the lights were off and the object was positioned to the right. In this scenario, YOLOv8 failed to detect faces, whereas YOLOv5 still detected them, though the recognition results showed an "unknown" classification. Notably, YOLOv8 performed better when the face was positioned on the right side under normal lighting conditions (lights on). Overall, the YOLOv5x model achieved an average accuracy of 88.4%, correctly recognizing 15 out of 18 tested faces. Similarly, YOLOv8x achieved an accuracy of 82.2%. In the online testing results, face detection and recognition for two

faces aligned in rows at the left, middle, and right positions under normal and dark conditions are shown in Table 5. The YOLOv5x model achieved an average accuracy of 91.5%, correctly recognizing 19 out of 24 tested faces, while the YOLOv8x model obtained an average accuracy of 79.2%, correctly recognizing 16 out of 24 tested faces. These results indicate that the YOLOv5x model is more robust at detecting and recognizing two faces under poor lighting conditions and when the faces are positioned in the middle. This suggests that light intensity, distance, and position significantly impact the results of face detection and recognition. Notably, the YOLOv5x model can still recognize two faces when they are positioned on the right side or at an angle of less than 60 degrees to the camera, highlighting its robustness across various scenarios.

The online testing results for face detection and recognition of three faces lined up in rows in the left, middle, and right positions under normal and dark conditions are shown in Table 6. The YOLOv5x model achieved an average accuracy of 83.7%, correctly recognizing 22 out of 30 tested faces, while the YOLOv8x model obtained an average accuracy of 81.5%, correctly recognizing 20 out of 30 tested faces.

These results indicate that real-time testing for a single face position has little impact on recognition accuracy; the model can still accurately recognize multiple faces. In both normal and dark conditions, the results show that the model performs similarly when testing with one person, as seen in Table 4. However, when dealing with two or three faces, the model has more difficulty recognizing all faces correctly. This is because processing time increases significantly as the number of faces to detect and recognize increases. The model needs to evaluate more areas of the image and attempt to detect additional objects, which can reduce detection speed. Moreover, the model may also misidentify faces, as seen in Table 6, number 6, and Figure 13, where YOLOv8x incorrectly identifies Sample2's face as Sample3's due to similar features being detected for a few seconds. Interestingly, distance plays a significant role in face recognition: the closer the face is, the easier it is for the model to recognize it, as facial features are more visible at close range than at a distance.

In addition, this study includes a condition labeled as 'unknown.' The 'unknown' condition is triggered when a face object has a confidence probability or accuracy below 0.70. Figure 14 illustrates an example of test results

Table 6. The results of online testing under conditions of light on and off with three face objects to be detected.

No	Condition	Sample	YOLOv5x	YOLOv8x
1	Light on – 3 objects in the left lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Undetected
		Sample 3	Recognized	Undetected
2	Light on – 3 objects in the left in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Unknown	Recognized
3	Light off –3 objects in the left lined up	Sample 1	Unknown	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Misrecognized	Recognized
4	Light off –3 objects on the left in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Undetected	Undetected
5	Light on – 3 objects in the center lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Recognized	Undetected
6	Light on – 3 objects in the center in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Recognized	Misrecognized
7	Light off – 3 objects in the center lined up	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Misrecognized	Unknown
8	Light off – 3 objects in the center in rows	Sample 1	Recognized	Recognized
		Sample 2	Recognized	Recognized
		Sample 3	Recognized	Undetected
9	Light off – 3 objects in the right lined up	Sample 1	Undetected	Undetected
		Sample 2	Recognized	Misrecognized
		Sample 3	Undetected	Undetected
10	Light off – 3 objects in the right in rows	Sample 1	Recognized	Recognized
		Sample 2	Undetected	Recognized
		Sample 3	Undetected	Undetected
11	Light on – 3 objects in the right in rows	Sample 1	Undetected	Undetected
		Sample 2	Recognized	Recognized
		Sample 3	Undetected	Undetected
12	Light off – 3 objects in the right line up	Sample 1	Undetected	Undetected
		Sample 2	Undetected	Recognized
		Sample 3	Undetected	Undetected

with unrecognized face objects. Since this face is not included in the dataset, the security system can effectively distinguish it as unrecognized. A confidence threshold of 0.70 was chosen as the cutoff between recognized and unrecognized individuals because this research prioritizes detecting facial objects for a security system that requires low false-positive rates. By using a confidence threshold of 0.70, we can reduce false positives by predicting only with high confidence scores. Additionally, a high confidence threshold can decrease the number of predictions the model makes, thereby reducing the computational load required to evaluate images and draw bounding boxes around objects.

4. Conclusion

Based on the research findings, YOLOv5x achieved the best results with an mAP of 83%, followed closely by YOLOv8x with an accuracy of 85.2%. In offline testing, all models correctly recognized faces across all classes, though their accuracy and confidence varied. Specifically, YOLOv5m achieved an average accuracy of 94%, YOLOv5x 95%, YOLOv8m 90%, and YOLOv8x 91%. In online testing, the YOLOv5 model consistently recognized more faces across all tested locations. Notably, lighting conditions significantly affected face recognition, with better results under normal lighting than under darker conditions.

Furthermore, the number of individuals present in a single frame had a profound impact on the model's performance. As the number of faces in a frame increased, the model struggled to accurately detect and recognize them. This is because the model requires more computational resources to process frames with multiple face classes, making it more challenging to achieve high accuracy.

The research findings also demonstrated that the proposed method could detect and recognize faces at distances of 1-3 meters from the CCTV camera, even when subjects' faces were positioned on the right side or not directly facing the lens. Specifically, the YOLOv5x model achieved an average detection and recognition precision accuracy of 87.8%. This research has also demonstrated its potential to enhance indoor security systems by distinguishing between known and unknown individuals in the database. When a face is detected in a CCTV frame with a confidence score below 0.70, it is classified as "Unknown", enabling effective identification of unfamiliar individuals.

In the future, it is necessary to investigate a more diverse range of face orientations and expressions, especially when faces are not directed toward the camera or when multiple faces appear in a crowd. Additionally, it is important to study face occlusions, crowd conditions, and deep models such as transformers.

Figure 13. The results of misrecognition under the condition of light on with three objects in the center in rows.

Figure 14. The results of unknown.

Acknowledgements

The authors would like to express their sincere gratitude to Universitas Sriwijaya for providing a supportive research environment and academic atmosphere that enabled the completion of this study. The authors also gratefully acknowledge financial support from Universitas Sriwijaya through its research funding program.

Funding

This study was funded by a DIPA of the Public Service Agency of Universitas Sriwijaya (No. SP DIPA-023.17.2.677515/2023), in accordance with the Rector's Decree No. 0118/UN9.3.1/SK/2023, dated April 18, 2023.

Conflict of Interest

The authors declare no conflicts of interest.

References

Abhinand, A., Mulerikkal, J., Antony, A., Aparna, P. A., & Jaison, A. C. (2021). Detection of moving objects in a metro rail CCTV video using YOLO object detection models. In *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2021, Volume 1* (pp. 183-195). Singapore: Springer Singapore.
https://doi.org/10.1007/978-981-16-2934-1_12

Aung, H., Bobkov, A. V., & Tun, N. L. (2021). Face detection in real time live video using yolo algorithm based on Vgg16 convolutional neural network. In *2021 International conference on industrial engineering, applications and manufacturing (ICIEAM)* (pp. 697-702). IEEE.
https://doi.org/10.1109/ICIEAM51226.2021.9446291

Casas, E., Ramos, L., Bendek, E., & Rivas-Echeverría, F. (2024). Yolov5 vs. yolov8: Performance benchmarking in wildfire and smoke detection scenarios. *Journal of Image and Graphics*, *12*(2), 127-136.
https://www.joig.net/2024/JOIG-V12N2-127.pdf

Dwyer, B., Nelson, J., Hansen, T., et. al., Roboflow (Version 1.0) Retrieved Jul. 03, 2023, from
https://roboflow.com

Fahad, S., ur Rahman, S., Khan, I., & Haq, S. (2017). An experimental evaluation of different face recognition algorithms using closed circuit Television images. In *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)* (pp. 51-54). IEEE.
https://doi.org/10.1109/SIPROCESS.2017.8124504

Guo, S., Li, L., Guo, T., Cao, Y., & Li, Y. (2022). Research on mask-wearing detection algorithm based on improved YOLOv5. *Sensors*, *22*(13), 4933.
https://doi.org/10.3390/s22134933

Halawa, L. J., Wibowo, A., & Ernawan, F. (2019). Face recognition using faster R-CNN with inception-V2 architecture for CCTV camera. In *2019 3rd international conference on informatics and computational sciences (ICICoS)* (pp. 1-6). IEEE.
https://doi.org/10.1109/ICICoS48119.2019.8982383

Kanyal, H. S., Goel, M., Tomar, A. S., Yadav, H. K., & Singh, K. (2020). Object recognition and security improvement by enhancing the features of CCTV. In *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)* (pp. 245-248). IEEE.
https://doi.org/10.1109/SMART50582.2020.9337065

Khalfaoui, A., Badri, A., & Mourabit, I. E. (2022). Comparative study of YOLOv3 and YOLOv5's performances for real-time person detection. In *2022, 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (pp. 1-5). IEEE.
https://doi.org/10.1109/IRASET52964.2022.9737924

Majeed, F., Khan, F. Z., Nazir, M., Iqbal, Z., Alhaisoni, M., Tariq, U., ... & Kadry, S. (2022). Investigating the efficiency of deep learning based security system in a real-time environment using YOLOv5. *Sustainable Energy Technologies and Assessments*, *53*, 102603.
https://doi.org/10.1016/j.seta.2022.102603

Menaka, K., & Yogameena, B. (2021). Face detection in blurred surveillance videos for crime investigation. In *Journal of Physics: Conference Series* (Vol. 1917, No. 1, p. 012024). IOP Publishing.
https://doi.org/10.1088/1742-6596/1917/1/012024

Mun, H. J., & Lee, M. H. (2022). Design for visitor authentication based on face recognition technology using CCTV. *IEEE Access*, *10*, 124604-124618.
https://doi.org/10.1109/ACCESS.2022.3223374

Nurhopipah, A., & Harjoko, A. (2018). Motion detection and face recognition for cctv surveillance system. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 12(2), 107-118.
https://doi.org/10.22146/ijccs.18198

Obaida, T. H., Flaih, N., & Jamil, A. S. (2022). Comparative of Viola-Jones and YOLO v3 for Face Detection in Real time. *Iraqi Journal Of Computers, Communications, Control And Systems Engineering*, *22*(2), 6.
https://ijccce.uotechnology.edu.iq/journal/vol22/iss2/6/

Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics*, *2*(1), 1-12.
https://doi.org/10.33736/jcsi.5070.2023

Qi, D., Tan, W., Yao, Q., & Liu, J. (2022). YOLO5Face: Why reinventing a face detector. In *European Conference on Computer Vision* (pp. 228-244). Cham: Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-25072-9_15

Sholahuddin, M. R., Harika, M., Awaludin, I., Dewi, Y. C., Fauzan, F. D., Sudimulya, B. P., & Widarta, V. P. (2023). Optimizing yolov8 for real-time cctv surveillance: A trade-off between speed and accuracy. *Jurnal Online Informatika*, *8*(2), 261-270.
https://doi.org/10.15575/join.v8i2.1196

Sino, H. W., & Areni, I. S. (2019). Face Recognition of Low-Resolution Video Using Gabor Filter & Adaptive Histogram

Equalization. In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)* (pp. 417-421). IEEE.
https://doi.org/10.1109/ICAIIT.2019.8834558

Son, N. T., Anh, B. N., Ban, T. Q., Chi, L. P., Chien, B. D., Hoa, D. X., ... & Hassan Raza Khan, M. (2020). Implementing CCTV-based attendance taking support system using deep face recognition: A case study at FPT polytechnic college. *Symmetry*, *12*(2), 307.
https://doi.org/10.3390/sym12020307

Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine learning and knowledge extraction*, *5*(4), 1680-1716.
https://doi.org/10.3390/make5040083

Ullah, R., Hayat, H., Siddiqui, A. A., Siddiqui, U. A., Khan, J., Ullah, F., ... & Karami, G. M. (2022). A real-time framework for human face detection and recognition in CCTV images. *Mathematical Problems in Engineering*, *2022*(1), 3276704.
https://doi.org/10.1155/2022/3276704

Wang, G., Ding, H., Duan, M., Pu, Y., Yang, Z., & Li, H. (2023). Fighting against terrorism: A real-time CCTV autonomous weapons detection based on improved YOLO v4. *Digital Signal Processing*, *132*, 103790.
https://doi.org/10.1016/j.dsp.2022.103790

Xu, Q., Zhu, Z., Ge, H., Zhang, Z., & Zang, X. (2021). Effective face detector based on YOLOv5 and superresolution reconstruction. *Computational and mathematical methods in medicine*, *2021*(1), 7748350.
https://doi.org/10.1155/2021/7748350