INTERNATIONAL JOURNAL OF APPLIED ARTIFICIAL INTELLIGENCE AND ROBOTICS
E-ISSN: 3124-1212, DOI: 00.000/XXXX.YYYY.II.NNN
Vol. 2 No. 1 2026 pp. 22-28

# An Intelligent Student Attendance System Based on Facial Image Recognition Using YOLOv5: A Case Study at Politeknik Negeri Lhokseumawe

Heri Maulana[1], Azhar[2], Rahmad Hidayat[3], Eni Mawardhaningrum[4*]

[1,2,3] Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Indonesia
[4] Sekolah Pascasarjana, Universitas Muhammadiyah Yogyakarta, Indonesia

*Corresponding Author: eny_0909@ mail.umy.ac.id

**Article info:** Received 02-02-2026, Revised 19-02-2026, Accepted 15-03-2026



## Abstract

You Only Look Once (YOLO) is an effective deep learning method for real-time object detection. This method is applied to build an accurate, fast facial recognition model, with a case study on an attendance system based on facial images at the Lhokseumawe State Polytechnic. The main objective is to implement and evaluate the YOLOv5s model's performance on the attendance system using facial images with high accuracy. The research process includes collecting facial image datasets from 40 subjects with various viewing angles to train the YOLOv5s model. This model is specifically configured to detect one class of objects, namely faces, and then integrated into the system to function as the main face detector. Model performance is evaluated quantitatively using a confusion matrix to measure key metrics such as accuracy, precision, recall, and F1 score. The evaluation results show that the developed YOLOv5s model has excellent performance. This model achieved 90% accuracy, with a precision and recall of 92%. The balanced F1-score value (92%) proves that the YOLOv5s method has a high level of accuracy for detecting faces. The high-performance metrics confirm that this method is the right solution for building an attendance system based on accurate facial images.

**Keywords:** attendance,student, recognition, face, image, YOLO5s

## 1. Introduction

The development of information and communication technology has encouraged the implementation of digital systems in various aspects of education, including student attendance. In many universities, including the Lhokseumawe State Polytechnic (PNL), QR code-based attendance systems have been used as an alternative to manual attendance due to their practicality and efficiency. However, this system has a fundamental weakness: the high potential for proxy attendance, where students who are not physically present can still be recorded as present by using QR codes shared by other students [1]. This situation negatively impacts academic integrity, student discipline, and the overall quality of the learning process. To address this issue, an attendance system is needed that can verify student attendance directly and authentically. One widely developed approach is biometric-based attendance, specifically facial recognition. The face is a unique biometric characteristic for each individual and is difficult to forge, making it more secure than card-based or digital code-based methods. Furthermore, facial recognition systems are non-intrusive and can be integrated with relatively easy-to-find campus cameras [2].

The success of a facial recognition system is greatly influenced by the facial detection method used. Pre-deep learning face detection methods such as Haar Cascade, Histogram of Oriented Gradients (HOG), and Support Vector Machine (SVM) have been widely used in early research [3][4]. Although these methods have low computational complexity, their performance tends to degrade under unstable lighting conditions, variations in facial angles, facial expressions, and complex backgrounds [5]. This results in inconsistent attendance system accuracy when applied to dynamic classroom environments. Furthermore, two-stage detection methods such as R-CNN, Fast R-CNN, and Faster R-CNN offer higher accuracy than classical methods. However, these approaches require two main stages: region proposal and classification, requiring longer computation time and greater hardware resources. These limitations make two-stage methods less optimal for real-time applications, especially in attendance systems that must process many faces simultaneously in a short time [6].

Alternatively, the You Only Look Once (YOLO) algorithm is a single-stage object detection method that combines object detection and classification in a single computational step [7]. YOLO processes the entire input image end-to-end in a single forward pass through the neural network, resulting in significantly higher inference speeds than two-stage methods. This advantage makes YOLO highly effective for real-time applications, including facial recognition-based attendance systems. YOLO divides the input image into multiple grids and simultaneously predicts bounding boxes, confidence scores, and object classes [8]. With this approach, YOLO can detect multiple objects in a single image with a short processing time. In the context of student attendance, this capability is crucial because a class can include dozens of students who must be detected and recognized simultaneously without disrupting the lecture [9][10].

Several recent studies have shown that combining the YOLO face detector with a deep learning-based face recognition method can significantly improve attendance system performance, both in terms of speed and accuracy. YOLO detects facial locations quickly and precisely, while the facial recognition module matches detected faces to a student database [11]. This approach can reduce detection errors and increase system reliability under varying environmental conditions. Along with the development of YOLO, various versions have emerged, starting from YOLOv3, YOLOv4, YOLOv5, and YOLOv8. Each version offers improvements in architecture, accuracy, and computational efficiency. In this study, YOLOv5s was chosen as the primary face detection model. YOLOv5s is a lightweight variant of YOLOv5 that offers a balance between speed and accuracy, making it suitable for implementation on hardware with moderate specifications [12]. In addition, YOLOv5 has comprehensive documentation, a large user community, and active development support, thus facilitating the system training and implementation process [13][14].

Although YOLOv8 is reported to have a slightly higher accuracy rate, this improvement is generally accompanied by greater computational requirements [15]. In the context of implementation at Lhokseumawe State Polytechnic, the selection of YOLOv5s is considered more relevant because it considers hardware availability, system efficiency, and real-time application needs. Tests show that YOLOv5s is capable of performing real-time face detection with an average processing time of around 4.1 milliseconds per image, and is capable of detecting and recognizing multiple faces simultaneously in a single frame [16]. YOLOv5s' ability to detect multiple faces simultaneously is very suitable for classroom conditions in a campus environment, where students have varying seating positions, distances, and viewing angles towards the camera [17]. With stable performance, the face-based attendance system can be effectively integrated into the lecture process, so that attendance can be carried out automatically, accurately, and efficiently.

## 2. Methods

This research applies a structured machine learning model development methodology, starting from data collection to model evaluation.

### 2.1 Dataset

This study utilized primary data comprising facial images collected from 40 students at Lhokseumawe State Polytechnic. Data was gathered by regularly photographing each participant using a smartphone camera, chosen for its widespread availability and ease of integration into future attendance systems. To ensure the dataset addressed real-world challenges, photos were captured in a variety of poses: front view, 45-degree side view, and 90-degree side view. Additional angles—including right, top, left, and bottom—were included to help the model become pose-invariant, enabling accurate recognition even when subjects were not facing the camera directly. Approximately 100 photographs were taken of each student to provide sufficient data for the algorithm to thoroughly learn individual features. All images were stored in JPEG format to balance storage efficiency, processing speed, and visual quality.

### 2.2 You Only Look Once (YOLO)

You Only Look Once (YOLO) is an intelligent neural network designed for real-time object detection. YOLO applies a single neural network to the entire image, which is then divided into regions (grids) [15]. Each region will predict a bounding box and the probability that an object is present within it. For each box in the bounding region, YOLO computes a probability and classifies whether the box contains an object. This method allows YOLO to detect objects quickly and accurately, making it one of the good algorithms for real-time object detection. The input image is divided into a fixed-size grid, where each cell predicts the objects within its area. This approach offers the advantage of handling multiple objects in a single image. YOLO has several advantages over classifier-oriented systems, as demonstrated across the entire image when tested with globally informed predictions [16].

### 2.3 Traning and Testing

The training and testing flow for the attendance model used in this study is shown in Figure 1. In the training phase, the system builds a facial reference database for the attendance process. Student facial images are collected and preprocessed (size normalization and lighting correction) to ensure consistent quality. The YOLO model is used to detect facial regions and crop them. Facial cropping is then extracted using the face_recognition module to generate facial encodings, unique numeric representations for each individual. The entire encoding is stored in the database as an attendance model, which serves as a reference in the next phase. In the testing phase, the system performs facial recognition during attendance. Student facial images are captured and preprocessed to ensure adequate quality. YOLO then detects facial locations and crops facial areas. These croppings are extracted by face_recognition to generate encodings, which are then compared with the encodings in the attendance model database from the training phase. If a match is achieved, the student's identity is verified, and attendance is recorded in the system.

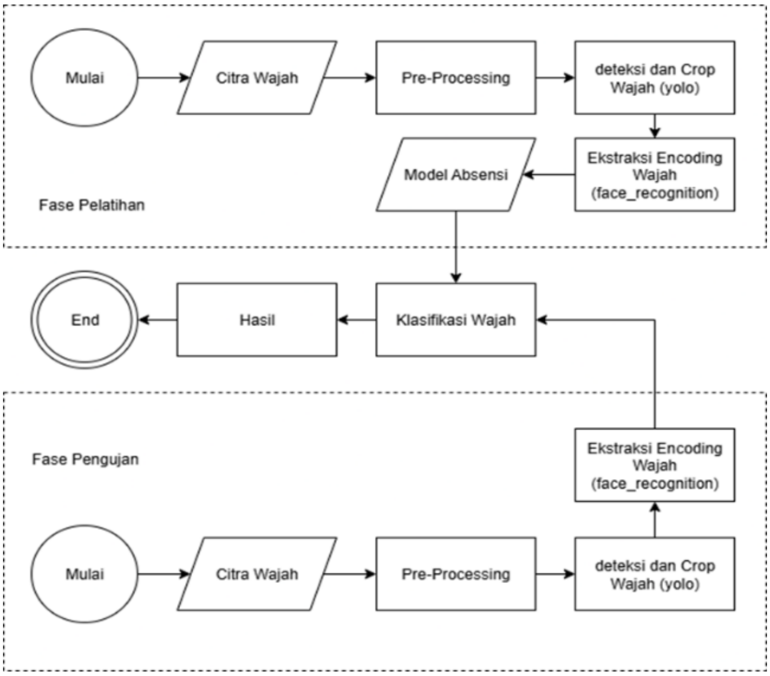


Figure 1. Model's Training and Testing Flow

Model evaluation is conducted to assess the performance of the attendance model in detecting and classifying objects (e.g., faces) on test data. The approach used is a confusion matrix to see the distribution of correct and incorrect predictions for each object. From the confusion matrix, the following main metrics are calculated: Accuracy (the proportion of all correct predictions to the total predictions), Precision (the accuracy of positive detections, i.e., the proportion of correct detections from all positive detections), Recall (the model's ability to find all objects that should be detected), and F1-score (the harmonic mean of Precision and Recall, which provides a balanced assessment when the class distribution is uneven).

The face attendance process in the system begins when the user opens the attendance menu. The camera is activated to capture a facial image at a distance of $\pm 10$ cm under normal lighting, without any covering accessories. The image is then processed by YOLO to detect the facial area and generate a bounding box along with a confidence score; the detected face is cropped and proceeds to the preprocessing stage and feature extraction using face_recognition to form a 128-dimensional vector encoding. This encoding is compared with the database using Euclidean distance; if the distance is below the threshold of 0.6, then the face is declared a match [17]. To prevent forgery, liveness verification (blinking/head movement) is performed using facial landmarks (dlib). If all stages are passed, the system records the attendance in the database and stores the facial image as evidence.

## 3. Result and Discussions

This section presents an analysis of the results from the attendance model training process, along with an evaluation of its performance using established testing metrics. The analysis was conducted to assess the model's ability to accurately recognize and record attendance, as well as to identify the effectiveness and reliability of the developed system.

### 3.1. Models Performance

The following are the performance results of the developed model shown in Figure 2 using a confusion matrix.

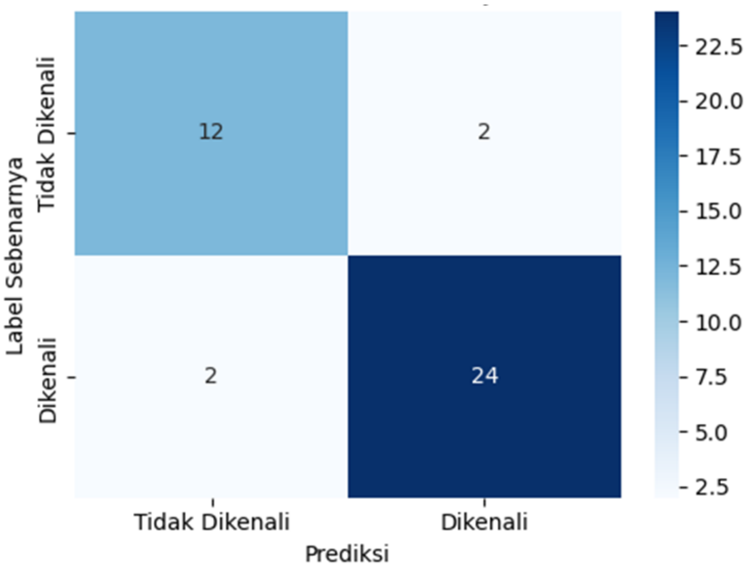


Figure 2. *Confusion Matrix* Result

The confusion matrix in this test uses the Y-axis for the actual label (Unrecognized/Recognized) and the X-axis for the predicted label (Unrecognized/Recognized), with each cell containing the number of samples. The results show TN = 12, namely unregistered faces predicted as unrecognized (correct); FP = 2, namely unregistered faces but predicted as recognized (false accept); FN = 2, namely registered faces but predicted as unrecognized (false reject); and TP = 24, namely registered faces predicted as recognized (correct). In total, there are 40 samples consisting of 14 unregistered faces and 26 registered faces which serve as the basis for calculating the evaluation metrics. To clarify how the model works, Figure 3 previously presented a confusion matrix that is then used as the basis for calculating and visualizing the evaluation metrics, namely accuracy, precision, recall, and F1-score. The matrix maps the actual labels (Unrecognized/Recognized) against the predicted output and produces four components: TN (12 cases of unregistered faces that were correctly rejected), FP (2 cases of unregistered faces that were incorrectly accepted), FN (2 cases of registered faces that were incorrectly rejected), and TP (24 cases of registered faces that were correctly accepted). In total, there are 40 samples: 14 unregistered and 26 registered, which form the basis for calculating these four metrics. The performance analysis in this section therefore, refers directly to the distribution of results in the matrix.

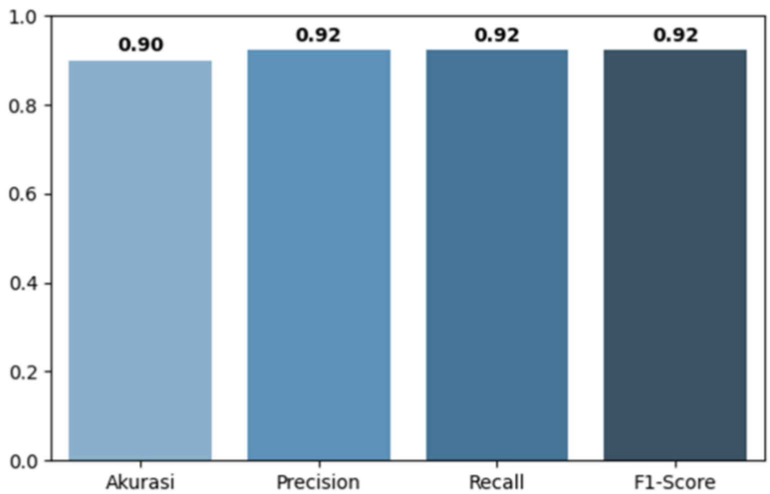


Figure 3. Confusion Matrix Results

Based on the data in Figure 3, the attendance system performs quite well. With an accuracy of 90%, this system has been proven to correctly classify most test data. Furthermore, the precision value (92%) highlights the system's ability to minimize identification errors (False Positives), indicating that the risk of incorrectly recognizing unregistered faces is very low. On the other hand, the recall value (92%) confirms that the system is very effective in recognizing almost all faces that should be registered, with a minimal detection failure rate (False Negatives). The F1-score value (92%) confirms a strong balance between precision and recall. Based on the evaluation results, the facial-image-based attendance system achieves 90% accuracy, 92% precision, 92% recall, and 92% F1-score. This strong balance between precision and recall indicates the system is able to minimize identification errors while accurately detecting student attendance.

## 3.2. User Interface Implementation

The following is a visualization of the attendance process that was successfully carried out by one of the students, which is shown in Figure 4.

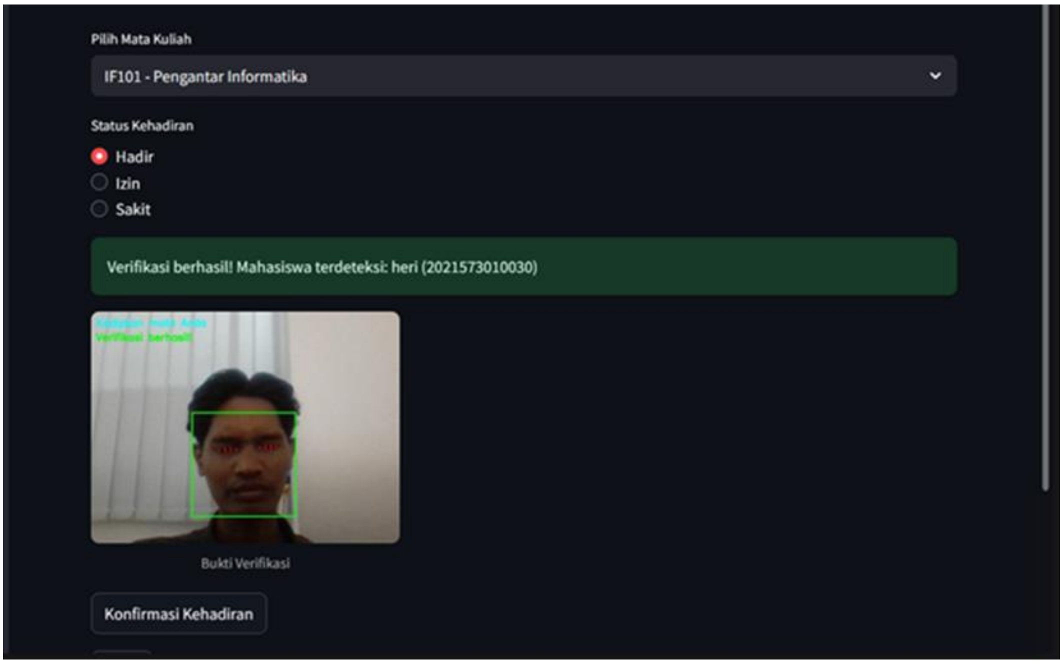


Figure 4. Student Attendance Visualization

The attendance visualization results in Figure 5 show the attendance system interface after a successful facial verification. The Verification Success status indicates that the user's identity has been validated, and a Confirm Attendance button is available to complete the attendance recording. This user-friendly interface provides clear feedback on verification status, so users can see that the facial recognition process has succeeded and the system is ready to record their attendance. Implementing the confirmation button as the final step serves as an additional security mechanism to prevent accidental or unwanted attendance recording, while also giving users the opportunity to cancel the process if necessary. The interface shown in the figure also displays additional information, including the name of the successfully identified user, the verification time, and the confidence level of the facial recognition process. The available and active Confirm Attendance button on the interface allows users to complete the attendance recording process manually, providing final control to users before the attendance data is officially recorded in the system database. The next visualization on Android is in Figure 6.

Figure 5 shows the interface of the Attendance System application based on facial recognition. There is a dropdown to select a course; in this example, it is IF101 – Database. The middle section displays a capture of the camera used to detect the student's face. Below are the attendance status and two control buttons: Activate Auto to enable automatic attendance, and Take Attendance for manual attendance. In the flow, the user selects a course → ensures the face is in the camera frame with adequate distance and lighting → selects Activate Auto for continuous scanning or Take Attendance for a single take → the application provides feedback on the results (success/failure, recognized name, time, and proof of a mini/thumbnail photo). Next, attendance data for students who have taken attendance is shown in Figure 6.

Figure 7 displays the attendance report page, which is used to retrieve and manage attendance data stored in the system. This page is equipped with various filters that make it easier for users to search for specific data, including a time range filter by selecting the start and end dates, a course filter to limit attendance data to specific courses, a student filter to retrieve attendance data for specific individuals, and an attendance status filter that allows data sorting based on the categories of attendance, permission, sick, or absent.


Figure 5. Interface on Android App

Once the filters are applied, the corresponding attendance data is displayed in a structured table and can be reviewed to ensure accuracy. Once the data meets your needs, users can easily download the report using the Download Report (CSV) button to obtain a CSV report or the Download Report (PDF) button to obtain a PDF report. This feature speeds up and enhances the report retrieval process.

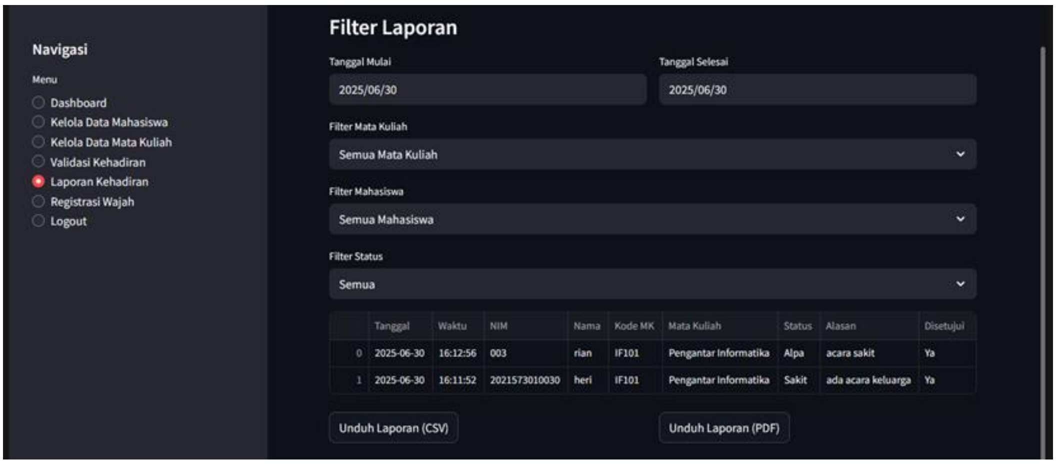


Figure 6. Attendance Data Results

# 4. Conclusion

Based on the tests carried out, it can be concluded that the system's functionality has been thoroughly tested using the black-box method, validating that all features, including login, facial registration, the attendance process, and report generation, function as expected. To evaluate the model performance, testing was carried out using a confusion matrix which showed very good results, with an accuracy level reaching 90% which means the system makes correct decisions in 90%, a precision of 92% indicates that of all faces identified as registered users, 92% of them are correct, a recall of 92% indicates that the system successfully recognizes 92% of the total users that should be recognized, and an f1-score of 92% is a balancing score and confirms that the model is very good at avoiding errors and in recognizing legitimate users.

# REFERENCES


[1] A. Nwabuwe, B. Sanghera, T. Alade, and F. Olajide, "Fraud Mitigation in Attendance Monitoring Systems using Dynamic QR Code, Geofencing and IMEI Technologies," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 4, pp. 938–945, 2023, doi: 10.14569/IJACSA.2023.01404104.

[2] M. S. Mohammed and K. A. Zidan, "Enhancing attendance tracking using animated QR codes: a case study," Indones. J. Electr. Eng. Comput. Sci., vol. 31, no. 3, pp. 1716–1723, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1716-1723.

[3] S. Joshi, S. Shinde, P. Shinde, N. Sagar, and S. Rathod, "Facial Recognition Attendance System using Machine Learning and Deep Learning," vol. 12, no. 04, pp. 152–157, 2023, [Online]. Available: www.ijert.org

[4] A. N and K. Anusudha, "Real time face recognition system based on YOLO and InsightFace," Multimed. Tools Appl., vol. 83, no. 11, pp. 31893–31910, Sep. 2024, doi: 10.1007/s11042-023-16831-7.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I–I. doi: 10.1109/CVPR.2001.990517.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] 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-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[8] 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," 2023. doi: 10.3390/make5040083.

[9] D. Sari and A. H. Mirza, "The Detection Of Face Recognition As Employee Attendance Presence Using The Yolo Algorithm (You Only Look Once)," J. Darma Agung, vol. 30, p. 41, Nov. 2022, doi: 10.46930/ojsuda.v30i3.2187.

[10] M.A Thalor and Omkar S. Gaikwad, "Facial Recognition Attendance Monitoring System using Deep Learning Techniques," Int. J. Integr. Sci. Technol., vol. 2, no. 1, pp. 45–52, 2024, doi: 10.59890/ijist.v2i1.1290.

[11] O. Faruque, F. Siddiqui, and S. Noor, "Face Recognition-Based Mass Attendance Using YOLOv5 and ArcFace," 2023, pp. 496–510. doi: 10.1007/978-3-031-34619-4_39.

[12] G. Jocher et al., "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022, Zenodo. doi: 10.5281/zenodo.7347926.

[13] Mardiana, M. A. Muhammad, and Y. Mulyani, "Library Attendance System using YOLOv5 Faces Recognition," in 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), 2021, pp. 68–72. doi: 10.1109/ICCTEIE54047.2021.9650628.

[14] I. N. G. A. Astawa, I. K. G. D. Putra, M. Sudarma, and R. S. Hartati, "KomNET: Face Image Dataset from Various Media for Face Recognition," Data Br., vol. 31, p. 105677, 2020, doi: https://doi.org/10.1016/j.dib.2020.105677.

[15] F. Indaryanto, A. Nugroho, and A. F. Suni, "Aplikasi Penghitung Jarak dan Jumlah Orang Berbasis YOLO Sebagai Protokol Kesehatan Covid-19," Edu Komputika J., vol. 8, no. 1, pp. 31–38, 2021, doi: 10.15294/edukomputika.v8i1.47837.

[16] F. Rachmawati and D. Widhyaestoeti, "Deteksi Jumlah Kendaraan di Jalur SSA Kota Bogor Menggunakan Algoritma Deep Learning YOLO," Pros. LPPM UIKA Bogor, pp. 360–370, 2020.

[17] H. Li and X. Zhu, "Face recognition technology research and implementation based on mobile phone system," in 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016, 2016, pp. 972–976. doi: 10.1109/FSKD.2016.7603310.