The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2025

Volume 33, Pages 96-104

IConTech 2025: International Conference on Technology

Enhancing Low-Resolution Facial Recognition in Classroom Environments Using YOLOv8

Gheri Febri Ananda
University of Gadjah Mada

Hanung Adi Nugroho
University of Gadjah Mada

Igi Ardiyanto
University of Gadjah Mada

Abstract: Accurate facial recognition is essential in modern classroom environments, enabling automated attendance tracking and real-time monitoring of student participation. However, classroom settings present unique challenges, including low-resolution images caused by distance, varied lighting conditions, and occlusions, which significantly reduce identification accuracy. While previous approaches often employed super-resolution methods to address these issues, they required high computational resources and offered suboptimal accuracy. This study proposes using YOLOv8 to enhance face detection and recognition specifically tailored for classroom conditions. Experiments were conducted with four YOLOv8 variants—YOLOv8-S, YOLOv8-M, YOLOv8-L, and YOLOv8-X—in real classroom settings involving 40 students within a 6 m x 5 m space. The results demonstrate that YOLOv8-X delivered the best performance, achieving 92% precision, 88% recall, and an mAP50 of 95%, proving highly effective for detecting students in challenging classroom scenarios. YOLOv8-L closely followed with 94% precision and 84% recall. In contrast, YOLOv8-M and YOLOv8-S showed limited effectiveness, with YOLOv8-S achieving only 82% precision and 70% recall. These findings highlight the suitability of YOLOv8-L and YOLOv8-X for addressing the complex challenges of classroom environments, providing robust solutions for improving facial recognition accuracy and efficiently automating classroom management systems.

Keywords: Low resolution, Face recognition, Yolov8

Introduction

Facial recognition in the classroom holds significant potential in supporting the concept of a smart classroom, particularly for automatic attendance tracking, monitoring student engagement, and personalizing learning interventions (Akash et al., 2023; Pabba & Kumar, 2022; Yin Albert et al., 2022). This technology enables teachers to track student attendance more efficiently, monitor student activities in real-time, and provide personalized learning interventions tailored to individual need. However, the classroom environment presents several challenges for facial recognition, such as variations in lighting, the distance between students and the camera, low image resolution, and occlusions, which can affect the accuracy of facial recognition (Gu et al., 2022; Shi & Tang, 2022). Therefore, addressing these challenges is essential for the system to be reliably implemented in educational settings.

One of the main problems in classroom settings is the low resolution of images caused by the distance between students and the camera, as well as the large number of individuals being monitored simultaneously. Numerous studies were conducted to address low-resolution challenges, focusing on techniques such as Super-Resolution

Convolutional Neural Networks (SRCNN) (Dong et al., 2016), Deep Convolutional Neural Networks (DCNN) (Horng et al., 2022), Super-Resolution Generative Adversarial Networks (SRGAN) (Zhao et al., 2023), and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) (Song et al., 2021). While these methods have shown promising results in improving image quality, These methods were primarily tested on general datasets and typically focus on recognizing individual faces rather than multiple faces within a classroom context. Additionally, these super-resolution methods require high computational power, which often poses a limitation when applied to devices with limited resources in educational settings. Most of these studies have yet to be tested under classroom conditions, where multitasking and simultaneous face recognition are crucial.

Figure 1. Illustration of challenges in the classroom environment

Various approaches have been developed to address the problem of face detection to support good performance in face recognition, particularly using enhanced multitask cascaded convolutional neural networks (MTCNN) and optimization of YOLOV3 with Bayesian (Gu et al., 2022; Shi & Tang, 2022). On the other hand, deep learning technology has also applied in facial recognition to improve the accuracy and efficiency of automatic identification processes (Khan et al., 2019; Nguyen et al., 2021). However, previous research still encountered issues with suboptimal accuracy. Additionally, several studies using YOLOv8 have successfully focused on small object detection in small images within remote sensing (Yue et al., 2024), manufacturing (Tao et al., 2023), and autonomous vehicles (Wang et al., 2024).

Therefore, this study aims to adopt YOLOv8 with the capability to recognize small objects and scale variations in classroom conditions. Furthermore, it seeks to provide a more effective and efficient solution to support smart classroom management. The study also aims to enhance system accuracy on a larger scale, making it applicable in classrooms with more students and diverse environmental conditions

Method

Dataset Collection

The dataset used in this study was collected from a real classroom environment, capturing various student orientations, lighting conditions, and distances from the camera. This dataset was specifically gathered to ensure that the designed system could be effectively adopted in real-world classroom settings. It consists of 181 images taken over multiple days, with 159 images used for training, 15 for validation, and 7 for testing. The images feature 40 students in different seating arrangements for each image. The dataset was captured using the classroom's existing camera, providing a realistic representation of typical classroom conditions. Figure 2 presents an example of the dataset used.

Figure 2. Example of image dataset used

To evaluate the system's performance on low-resolution facial recognition, face detection was carried out using the MTCNN method, followed by cropping the detected faces to analyze the resolution of individual faces across different seating positions. This approach aimed to demonstrate how face resolution fluctuates based on a student's proximity to the camera. Specifically, students seated in the front row exhibited face resolutions of approximately 40 x 52 pixels, those in the middle row had around 22 x 27 pixels, and those in the back row displayed face resolutions of about 16 x 21 pixels. Figure 3 provides examples of the cropped faces, illustrating the varying image resolutions captured from the dataset.

Figure 3. Face image resolution in the dataset (a) front row students, (b) middle row students, and (c), (d)back row students.

This data collection method highlights the challenges of recognizing all students in a classroom setting, particularly due to scale variation and low-resolution faces caused by distance from the camera. These differences present significant obstacles for facial recognition systems, emphasizing the need for models that can handle these variations to ensure robustness and effectiveness in real-world scenarios.

Architecture of Yolov8

The architecture of YOLOv8 builds upon the innovations introduced in its predecessors, integrating several advanced components designed to enhance performance in object detection tasks. It employs a custom backbone that features the C2f (Cross Stage Partial 2-Fusion) structure, which optimizes feature extraction by improving information flow between layers while reducing computational complexity (Tao et al., 2023). The neck incorporates SPPF (Spatial Pyramid Pooling Fast), allowing the model to effectively capture features at multiple scales, which is crucial for accurately detecting objects of varying sizes (Gunawan et al., 2023). The head of YOLOv8 introduces separate classification and detection heads, transitioning from an anchor-free to an anchor-based approach to improve localization and classification precision.

YOLOv8 also transitions from an anchor-free detection system to an anchor-based approach, which increases the precision of object localization and classification (Terven et al., 2023). The model further enhances performance by using Focal Loss for classification instead of Binary Cross-Entropy (BCE), and Distribution Focal Loss (DFL) for regression, improving bounding box predictions. Furthermore, the integration of Complete IoU (CIoU) loss refines the accuracy of bounding box dimensions, leading to higher detection precision (Yan et

al., 2023). These architectural improvements collectively result in significant gains in both accuracy and efficiency, making YOLOv8 particularly effective for real-time object detection applications. The architecture of YOLOv8 is illustrated in Figure 4.

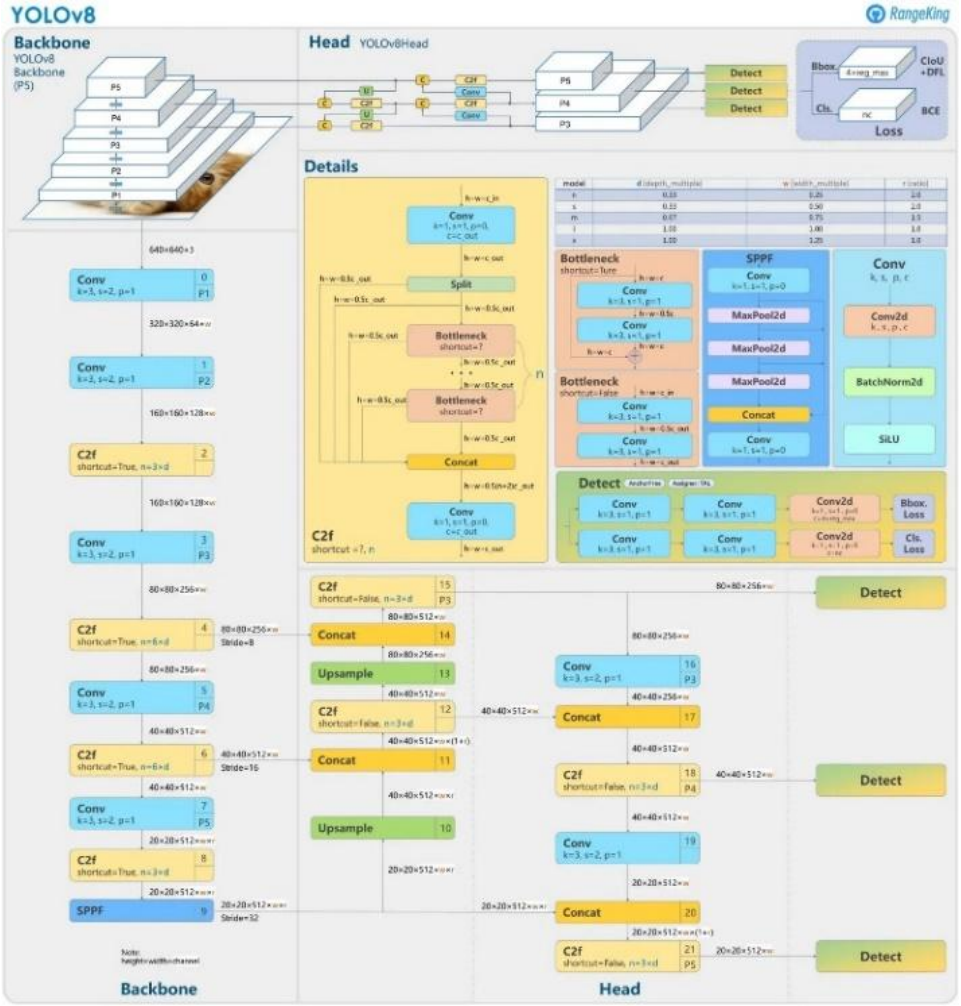

Figure 4. The architecture of YOLOv8 (Gunawan et al., 2023)

Model Selection and Training

In this study, several versions of YOLOv8 (YOLOv8-S, YOLOv8-M, YOLOv8-L, and YOLOv8-X) were trained and tested to determine the most effective model for face recognition in low-resolution classroom environments. The purpose of testing these different variants was to determine which version of YOLOv8 performed best in the specific case of low-resolution classroom environments. Each version was evaluated based on its ability to balance speed and accuracy, ensuring the optimal model for face detection in this setting was identified. Each variant presents different levels of complexity, characterized by the number of layers, parameters, and computational requirements (GFLOPs), as detailed in Table 1

Table 1. Yolov8 model variant details

Model Variant	Layers	Parameters	GFLOPs
YOLOv8-S	225	11.151.080	28.7
YOLOv8-M	295	25.879.480	79.2
YOLOv8-L	365	43.660.680	165.6
YOLOv8-X	365	68.191.128	258.3

To handle the computational demands of training, all models were trained for 100 epochs with a batch size of 16, using an image resolution of 640 x 640 pixels and an initial learning rate of 0.01. This training was conducted on Google Colab with an NVIDIA A100 GPU, providing the high processing power needed to

efficiently manage the models' computational requirements. The training configuration was standardized across all tests, as shown in Table 2.

Table 2. Hyperparameter setting

Hyperparameter	**Value**
Image Size	640 x 640
Epochs	100
Batch Size	16
Learning Rate	0.01
Processor Used	NVIDIA A100 (Google Colab)

These standardized hyperparameter settings were chosen to ensure consistent and fair comparison across the YOLOv8 variants.

Model Evaluation

After training, the performance of the YOLOv8 model is measured using several metrics, including precision, recall, and mean Average Precision (mAP). Precision is defined as the ratio of the number of true positive predictions (TP) to the total number of positive predictions made by the model, while Recall measures the ratio of the number of true positive predictions to the total number of actual positive objects present in the image, as expressed in Equations 1 and 2 below:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

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

Next, Average Precision (AP) is calculated as the area under the Precision-Recall curve, with the formula stated in Equation 3:

$$\text{AP} = \int_0^1 P\,(R)\,dR \tag{3}$$

This metric provides an overview of the model's performance at various threshold values. For further evaluation, mAP@0.5 is utilized to measure the accuracy of the model's detections at an Intersection over Union (IoU) threshold of 0.5. Additionally, mAP@0.5:0.95 calculates the average AP value across a range of IoU thresholds from 0.5 to 0.95, providing a comprehensive assessment of the model's performance. These metrics are expressed in Equations 4 and 5 :

$$\text{mAP}_{0.5} = \frac{1}{N} \sum_{i=1}^{N} AP_i \text{ , IoU} = 0.5 \tag{4}$$

$$\text{mAP}_{0.5:0.95} = \frac{1}{N} \sum_{i=1}^{N} AP_i \text{ , IoU} = 0.5 : 0.05 : 0.95 \tag{5}$$

These metrics collectively give a detailed evaluation of the model's ability to detect and localize objects accurately across varying levels of detection difficulty.

Results and Discussion

Training Process Results

To evaluate the effectiveness of the YOLOv8 models, we conducted a series of experiments to compare their performance across key metrics: precision, recall, mAP50, mAP50-95, and various loss values. The training was carried out on four model variants—YOLOv8-S, YOLOv8-M, YOLOv8-L, and YOLOv8-X—using the same dataset and hyperparameters to ensure consistency and fairness in the comparison. This comparative analysis aims to highlight the capability of each model in detecting and classifying objects, guiding the selection of the most appropriate variant for classroom scenarios. The training process results are illustrated in Figure 5 below.

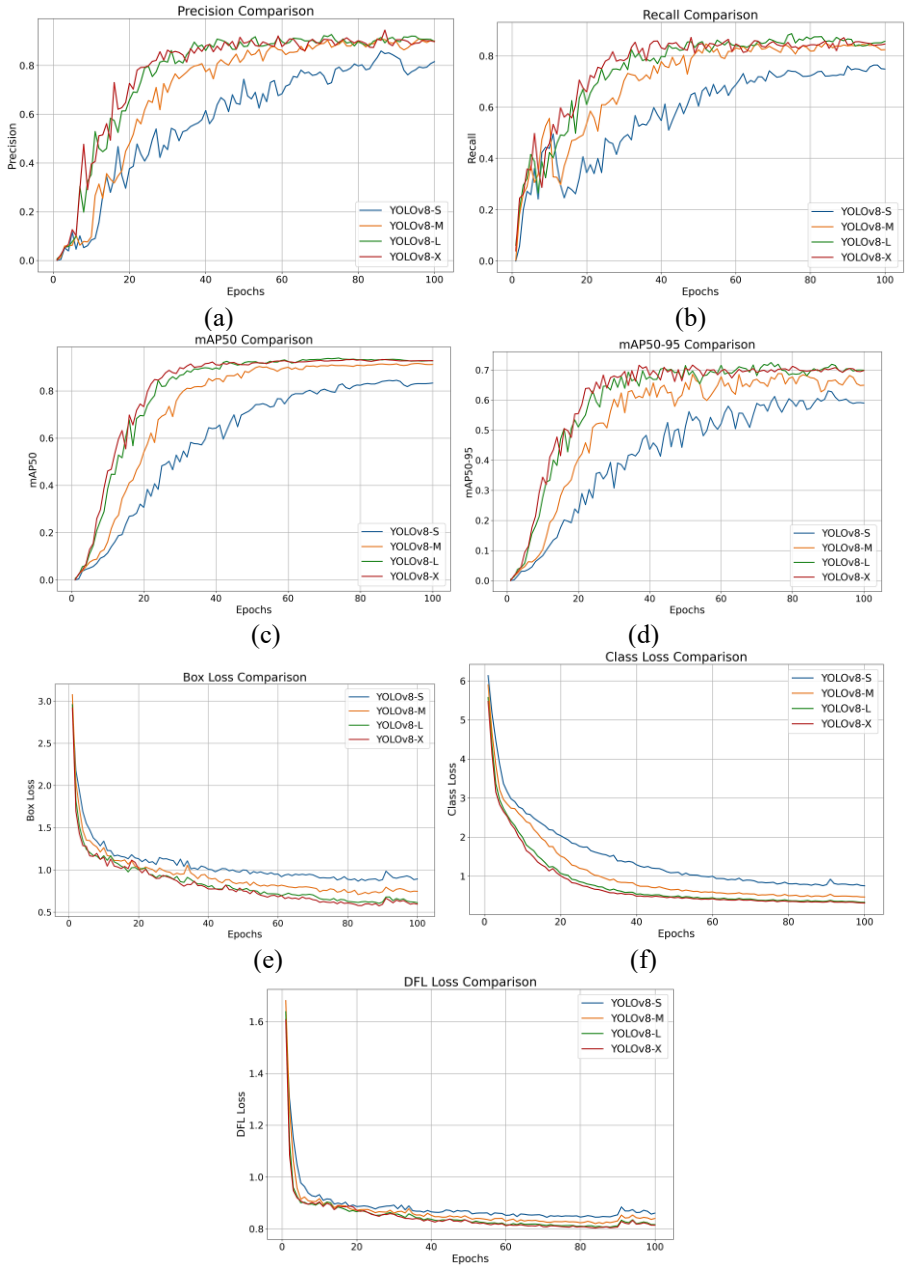

Figure 5. Evaluation metrics graph of four types of YOLOv8 during the training process (a) Precission, (b) Recall, (c) mAP50, (d) mAP50-9, (e) Box Loss, (f) Class Loss, (g) DFL Loss

During the training process, YOLOv8-L demonstrated the best stability in the recall metric, consistently achieving values in the 0.84-0.85 range over the last five epochs, indicating reliable and accurate object detection capabilities. YOLOv8-X also exhibited strong performance with high precision in the 0.89-0.91 range, although there were minor fluctuations in the recall metric compared to YOLOv8-L. On the other hand, YOLOv8-M had good precision around 0.91, but lower recall stability, with inconsistent values, reflecting poorer performance in detecting all objects in the images. Meanwhile, YOLOv8-S displayed the weakest performance, with precision around 0.82 and recall around 0.75, significantly lagging behind the other three models.

Training Process Results

To test the generalization capability of the model, an evaluation was conducted using test data that the model had never encountered during the training process. This testing aims to assess how well the model can recognize objects in previously unseen data and provides a more accurate representation of the model's performance in

real-world scenarios. The results of the evaluation are presented in the table 3, which lists key metrics such as precision, recall, mAP50, and mAP50-95 for each variant of the YOLOv8 model after training.

Table 3. Evaluation metric on four types of Yolov8

Yolov8 Types	Precission	Recall	mAP 50	mAP 50-95
Yolov8s	0.82	0.7	0.82	0.63
Yolov8m	0.91	0.83	0.91	0.71
Yolov8l	**0.94**	0.84	0.93	0.72
Yolov8x	0.92	**0.88**	**0.95**	**0.74**

Table 3 shows that YOLOv8-X outperforms with a precision of 0.92 and a recall of 0.88, accompanied by mAP50 of 0.95 and mAP50-95 of 0.74. This indicates that YOLOv8-X has excellent detection accuracy and can handle various object variations in the test data. YOLOv8-L also demonstrates strong performance with a precision of 0.94 and a recall of 0.84, as well as mAP50 of 0.93 and mAP50-95 of 0.72, reflecting the model's stability in detecting complex objects. In contrast, YOLOv8-M lags slightly behind with a precision of 0.91 and a recall of 0.83, along with an mAP50 of 0.91 and mAP50-95 of 0.71. This suggests that although YOLOv8-M has reasonably good accuracy, it may be less optimal in handling more complicated object variations compared to YOLOv8-L and YOLOv8-X. YOLOv8-S exhibits the lowest performance among the four models, with a precision of 0.82, recall of 0.70, mAP50 of 0.82, and mAP50-95 of 0.63, indicating that this model is less effective in detecting diverse objects in the test data.

Overall, the results from this testing reinforce the findings from the training phase, where YOLOv8-L and YOLOv8-X again demonstrate superior performance in object detection with high accuracy. These two models are more suitable for implementation in environments that require a high and diverse level of detection, such as the classroom scenarios being tested. Next, we present an overview of the tests performed with each model, aimed at deepening our understanding of their performance success in addressing the challenges of face recognition in demanding classroom environments. These examples are illustrated in Figure 6.

Figure 6. View on testing each type of YOLOv8 (a) YOLOv8-S, (b)YOLOv8-M, (c) YOLOv8-L, (d) YOLOv8-X

Figure 6 shows the effectiveness of the YOLOv8 model in recognizing faces under various challenging classroom conditions. These results highlight YOLOv8-L's robustness and adaptability, making it a valuable tool for applications in educational settings where reliable face recognition is essential. The high accuracy and

efficiency observed in these tests reinforce the model's potential for enhancing classroom management and student engagement through effective facial recognition technology.

Conclusion

Our experiments with various YOLOv8 models demonstrate that YOLOv8-L and YOLOv8-X are highly effective for face recognition in complex classroom environments. Both models exhibit high precision and recall, showcasing their ability to accurately identify and classify faces even in challenging real-world scenarios. YOLOv8-L stands out for its stability and reliability, while YOLOv8-X delivers superior precision, making it ideal for applications requiring extremely accurate results. In contrast, the limitations observed in YOLOv8-M and YOLOv8-S suggest that these models may not perform optimally in scenarios with significant variations in facial appearances. These findings emphasize the importance of selecting the most appropriate model based on the specific requirements of the operational context. Moreover, the success of YOLOv8 in addressing challenges with small or low-resolution faces in classrooms eliminates the need for super-resolution methods before face recognition. Consequently, adopting YOLOv8 not only improves accuracy but also reduces the computational load typically associated with super-resolution techniques.

Recommendations

Future research can focus on further improving YOLOv8's performance by training it with a larger and more diverse dataset that reflects a wider range of classroom scenarios. Integrating YOLOv8 with other image enhancement techniques could also enhance the model's accuracy and robustness, especially in challenging environments. Additionally, developing real-time face recognition systems for classroom settings will be critical for practical applications. These systems could include intuitive interfaces for teachers and administrators, enabling them to monitor and analyze student engagement more effectively. The implementation of this technology may hold the potential to create modern classrooms that are more interactive and innovative. By integrating face recognition systems with AI-based technologies, teachers may identify student engagement patterns in real-time, personalize teaching methods, and enhance the efficiency of the learning process.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM Journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* The writing and publication of this article were supported and funded by Lembaga Pengelola Dana Pendidikan (LPDP) Indonesia.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology (www.icontechno.net) held in Trabzon/Turkey on May 01-04, 2025.

References

Akash, S., Dinesh, V., Muthamil Selvan, S., Alfred Daniel, J., & Srikanth, R. (2023). Monitoring and analysis of students' live behaviour using machine learning. *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 98–103.

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307.

Gu, M., Liu, X., & Feng, J. (2022). Classroom face detection algorithm based on improved MTCNN. *Signal, Image and Video Processing*, *16*(5), 1355–1362.

Gunawan, F., Hwang, C.-L., & Cheng, Z.-E. (2023). ROI-YOLOv8-based far-distance face-recognition. *2023 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, 1–6.

Horng, S.-J., Supardi, J., Zhou, W., Lin, C.-T., & Jiang, B. (2022). Recognizing very small face images using convolution neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2103–2115.

Khan, M. Z., Harous, S., Hassan, S. U., Ghani Khan, M. U., Iqbal, R., & Mumtaz, S. (2019). Deep unified model for face recognition based on convolution neural network and edge computing. *IEEE Access*, *7*, 72622–72633.

Nguyen, D. D., Nguyen, X. H., Than, T. T., & Nguyen, M. S. (2021). Automated attendance system in the classroom using artificial intelligence and ınternet of things technology. *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, 531–536.

Pabba, C., & Kumar, P. (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems*, *39*(1), e12839.

Shi, D., & Tang, H. (2022). A new multiface target detection algorithm for students in class based on bayesian optimized YOLOv3 Model. *Journal of Electrical and Computer Engineering*, *2022*(1), 1–12.

Song, C., He, Z., Yu, Y., & Zhang, Z. (2021). Low resolution face recognition system based on ESRGAN. *3rd International Conference on Applied Machine Learning (ICAML)*, 76–79.

Tao, Q., Chen, Y., & Chen, H. (2023). A detection approach for wafer detect in ındustrial manufacturing based on YOLOv8. *2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, 1–6.

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.

Wang, H., Liu, C., Cai, Y., Chen, L., & Li, Y. (2024). YOLOv8-QSD: An ımproved small object detection algorithm for autonomous vehicles based on YOLOv8. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–16.

Yan, M., Fan, Y., Jiang, Y., & Fang, Z. (2023). A fast detection algorithm for surface defects of bare PCB based on YOLOv8. *2023 3rd International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, 559–564.

Yin Albert, C. C., Sun, Y., Li, G., Peng, J., Ran, F., Wang, Z., & Zhou, J. (2022). Identifying and monitoring students' classroom learning behavior based on multisource information. *Mobile Information Systems*, *2022*(1), 9903342.

Yue, M., Zhang, L., Zhang, Y., & Zhang, H. (2024). An ımproved YOLOv8 detector for multi-scale target detection in remote sensing ımages. *IEEE Access*, *12*, 114123–114136.

Zhao, Y., Yao, C., Li, X., & Shen, L. (2023). Face recognition using the ımproved SRGAN. *2023 8th International Conference on Image, Vision and Computing (ICIVC)*, 120–123.

Author Information

Gheri Febri Ananda
University of Gadjah Mada
Yogyakarta, Indonesia
Contact e-mail: *gherifebriananda1998@mail.ugm.ac.id*

Hanung Adi Nugroho
University of Gadjah Mada
Yogyakarta, Indonesia

Igi Ardiyanto
University of Gadjah Mada
Yogyakarta, Indonesia

To cite this article:

Ananda, G.F. Nugroho, H. A & Ardiyanto, I. (2025). Enhancing low-resolution facial recognition in classroom environments using YOLOv8. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 33,* 96-104.