Electronic Integrated Computer Algorithm Journal
Vol. 3, No. 1, October 2025, pp 17~23

ISSN: 3031-0350, DOI: 10.62123/enigma.v3i1.107

Application of Region of Interest (ROI) in Student Attendance Detection System in Classroom

Setyo Fahmi Noor Faizi[1], Al-Khowarizmi[1*]

[1] *Department of Information Technology, Universitas Muhammadiyah Sumatera Utara*, Indonesia.

DOI : 10.62123/enigma.v3i1.107

Received : September 09, 2025
Revised : October 05, 2025
Accepted : October 12, 2025

Keywords:
Attendance Detection, People Counting, ROI , Image Processing, HOG, RSM, Haar Cascade, Monitoring System

ABSTRACT

Efficient classroom management is a crucial requirement in academic environments such as the Faculty of Computer Science and Information Technology to increase productivity. This study aims to design and evaluate a real-time presence detection and counting system by implementing the Region of Interest (ROI) method to improve computational efficiency and accuracy. This methodology involves the use of a Logitech C270 HD webcam, with a static ROI set at 90% of the central video frame to focus the analysis. Person detection and counting are performed using a combination of Histogram of Oriented Gradients (HOG) for the body and Haar Cascade for the face. Time series reasoning with a minimum duration of 60 seconds and a grace period of 5 seconds is implemented to validate presence and stabilize the room status, with system performance evaluated using Precision and Recall metrics. The results show that the system successfully displays the status and number of people in the room very well, but the evaluation shows a Recall value of 1.00, which means the system detects every actual human presence. However, this system has significant accuracy issues, indicated by a low Precision of 0.04 and a high number of False Positives of 710. In conclusion, although the ROI application successfully improves the computational load and the temporal logic stabilizes the output, the HOG and Haar Cascade models are inadequate to handle visual noise in the ROI, resulting in low Precision and indicating the need for more sophisticated detection models.

1. INTRODUCTION

Digital image-based technology has become an important solution in indoor surveillance systems [1]. A digital image is a two-dimensional array with its f(x,y) values converted into discrete form for both image coordinates and brightness [2]. Image processing can generally be defined as the digital processing of a two-dimensional image [3]. Image processing is the process of processing and analyzing images that heavily involves visual perception [4]. This process has input data and output information in the form of images [5].

One of the methods that is widely used is Region of Interest (ROI)[6], ROI is an image segmentation [7] process in DIP (Digital Image Processing) which aims to separate or remove the background, so that the foreground or object of observation is easy to analyze [8]. In this detection, the Haar Cascade [9] and HOG methods are also used to assist in the detection [10][11], Haar cascade is used for face detection in a process consisting of several stages, namely feature extraction, classifier formation, and face detection [12] The hog method itself is used to detect human body shape. Each image has characteristics indicated by the gradient distribution. These characteristics are obtained by dividing the image into small regions called cells. Each cell is composed of a histogram of a gradient, and the combination of these gradients is used as a descriptor that represents an object [13].

In this study, ROI was applied to classroom monitoring at Faculty of Computer Science and Information Technology. The system was designed to focus on important areas within the room, such as doors, windows, or walls, that have distinctive patterns or structures. Furthermore, the Haar Cascade and HOG methods were used to detect other objects that might enter the camera's field of view, such as human presence. Thus, this system not only assists in room management but also improves the safety and comfort of room users. In developing this research, the researchers used Python because Python is claimed to combine capabilities and capabilities with very clear code syntax, and is equipped with a large and comprehensive standard library functionality [14].

In academic environments, such as the Faculty of Computer Science and Information Technology (FIKTI), automated classroom management is a crucial need. Efficiently utilized classrooms can improve productivity and enhance the comfort of teaching and learning. By implementing ROI-based technology, supported by the Haar Cascade and HOG methods [10], the system can detect human presence and provide real-time information about room usage [15]. For example, if a person is detected in a room for more than 1 minute, the system can send a notification to the room management database. This can help make room management more efficient and organized [16][17].

*Corresponding Author Email: alkhowarizmi@umsu.ac.id

2. METHODOLOGY

The research methodology used in this study uses Region of Interest (ROI). ROI is a segmentation technique used in image processing [18], allowing users to process images containing desired image data information. ROI works by encoding specific areas of a digital image differently, ensuring that more important areas have better quality than the surrounding areas [19]. This process can be seen in the research stages, outlined in the general architecture shown in Figure 1.

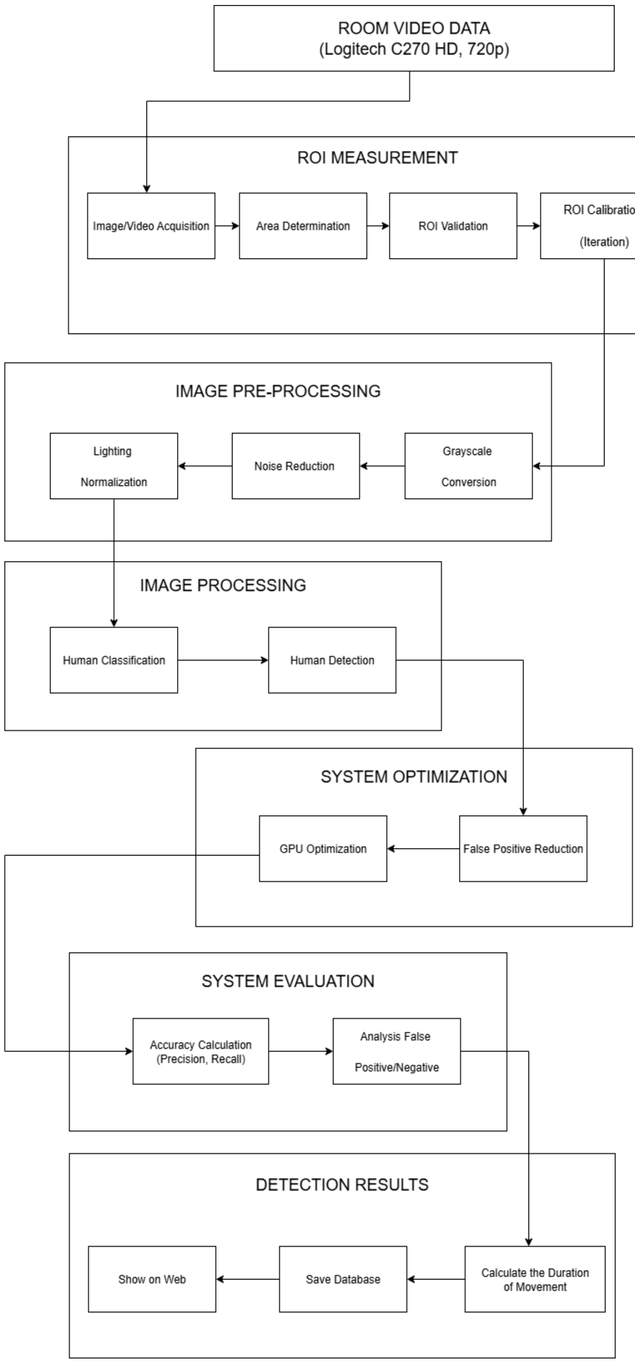

Figure 1. General Architecture

2.1 ROI Measurement

ROI measurement aims to determine specific areas within a video frame that are the primary focus of analysis. This process begins with image or video acquisition, which involves capturing frames from raw video for further processing. Next, the area is determined by manually selecting an ROI, such as a door or window, or automatically, such as using edge detection. Once the area is determined, ROI validation is performed to ensure that the selected area encompasses relevant objects, such as human movement paths. The final stage involves iterative ROI calibration [20], which involves readjusting the ROI boundaries based on the initial detection results to avoid interference from noise or unnecessary areas.

2.2 Image Processing

In the image processing stage, analysis is performed to extract important information from the pre-processed image. This process includes human detection, which identifies the location of humans within each video frame, and human classification to ensure that detected objects are truly human and not animals or other similar objects.

2.3 System Evaluation

During the system evaluation phase, performance measurements are conducted to assess the effectiveness of the method used. This evaluation includes accuracy using metrics such as precision, the ratio of correct detections to total detections, and recall, the ratio of correct detections to the number of actual cases. Furthermore, false positives and false negatives are analyzed to identify the causes of detection errors, such as poor lighting conditions or the presence of objects that resemble humans.

2.4 Detection Results

The detection results are the final output of the system, containing various important information. The system calculates the duration of movement activity by using the timestamp or frame rate of the video to determine how long a person has been in the monitoring area. These detection results are then displayed in real time via a web interface.

3. RESULTS AND DISCUSSION

3.1 Room Video Data

In the test, video data was captured live with a Logitech C270 HD webcam at 720p (1280×720 pixels). Data was captured under varying room conditions to test whether the program still functions in rooms with different structures.

Figure 2. Video Capture Screenshot

3.2 ROI Measurement

ROI measurement is implemented to focus detection on the most relevant areas. This approach effectively reduces computational load and minimizes interference from unimportant background areas. The system automatically calculates and limits the analysis area to the central 90% of the camera's field of view. The boundaries of this area are clearly visualized via a yellow box in the video window, providing immediate feedback on the active monitoring zone.

Python code snippet in ROI measurement stage

```
# The detection area (ROI) configuration is set to 90%
ROI_PERCENTAGE = 0.90

# Define and crop the ROI area of the frame
h, w = frame.shape[:2]
roi_size = int(min(h, w) * ROI_PERCENTAGE)
x, y = (w - roi_size) // 2, (h - roi_size) // 2
roi = frame[y:y+roi_size, x:x+roi_size]

# Draw a yellow box for ROI visualization
cv2.rectangle(display_frame, (x, y), (x+roi_size, y+roi_size), (0, 255, 255), 2)
cv2.putText(display_frame, "Area Deteksi (ROI)", (x, y - 15),
cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 255, 255), 1)
```

3.3 Image Pre-Processing

The preprocessing stage in this system focuses on converting the color space to grayscale. The image within the ROI is converted to grayscale before analysis. This step is crucial for optimizing the performance of the face detection algorithm (Haar Cascade), which works faster and more efficiently on images with a single color channel. Next, each frame cropped according to the ROI is converted to gray_roi. This grayscale version is then used for face detection, while body detection uses the ROI that remains in color.

The Code snippet to show Image Pre-Processing stage

```
# Convert colored ROI to Grayscale
gray_roi = cv2.cvtColor(roi, cv2.COLOR_BGR2GRAY)

# Using gray_roi as input for face detector
faces = face_cascade.detectMultiScale(gray_roi, scaleFactor=1.1, minNeighbors=5)
```

3.4 Image Pre-Processing

Here, the system analyzes images using two detection models: HOG (Histogram of Oriented Gradients) for bodies and Haar Cascade for faces. The results not only determine presence but also count people using intelligent heuristics, providing a visualization of the detection box for each object found. The system no longer simply detects presence (yes/no). After both detection models are run, the number of people is estimated by taking the larger of the number of detected bodies and faces to avoid double counting. The system then draws red (for bodies) and green (for faces) boxes in real time to provide visual evidence of each detection.

The code snippet shows the stages of Image Processing

```
# Running both detection models
humans, _ = hog.detectMultiScale(roi)
faces = face_cascade.detectMultiScale(gray_roi, scaleFactor=1.1, minNeighbors=5)

# The logic of counting people by taking the maximum value
person_count = max(len(humans), len(faces))

# Draw a red box for each "Human" detection.
for (hx, hy, hw, hh) in humans:
    cv2.rectangle(display_frame, (x+hx, y+hy), (x+hx+hw, y+hy+hh), (0, 0, 255), 2)

# Draw a green box for each "Face" detection.
for (fx, fy, fw, fh) in faces:
    cv2.rectangle(display_frame, (x+fx, y+fy), (x+fx+fw, y+fy+fh), (0, 255, 0), 2)
```

3.5 System Optimization

Optimizations within this system focus on time-based reasoning to filter detection results, ensuring the displayed status is stable and only triggered by significant activity, not momentary movement. A detection must persist consistently for 60 seconds to change the room status to "TERISI," effectively reducing false positives. The system will allow a 5-second tolerance if the target disappears from view, preventing "flickering" statuses and improving accuracy. The code snippet shows the System Optimization stages

```
# Configure time threshold for optimization
MIN_DURATION = 60
GRACE_PERIOD = 5

# Optimization logic in the main loop
if human_detected:
    last_detection_time = current_time
    if presence_start is None: presence_start = current_time
elif last_detection_time and (current_time - last_detection_time >= GRACE_PERIOD):
    presence_start = None

# The "Terisi" status is only active if the detection duration is met.
new_status = 1 if presence_start and (current_time - presence_start) >= MIN_DURATION else 0
```

3.6 System Evolution

A separate script, evaluate.py, is provided for quantitative and automated system performance evaluation. This script measures how accurate the detection logic is when tested on a labeled video dataset. By processing test videos, the script compares each system prediction to the actual state (ground truth) to calculate industry-standard metrics such as precision and recall. The script also automatically stores frames where detection errors occur to facilitate qualitative analysis.

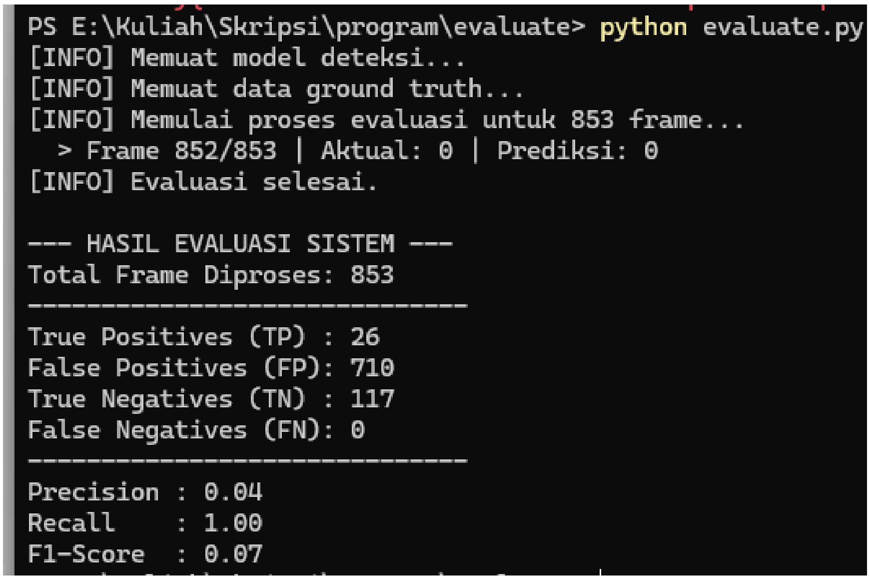

Figure 3. Evaluation Results

The evaluation results show a recall value of 1.00 and false negatives of 0, meaning your system successfully identified every frame that should have contained a human. This is an excellent achievement in terms of sensitivity. However, the very low precision value (0.04 or 4%) and False Positives (FP) of 710 are the main problems. This means that of all the "human" detections the system made, only 4% were truly accurate. Your system mistakenly assumed a human was present in an empty frame 710 times. This is likely due to specific conditions in the video, such as lighting, camera angle, or objects in the background, that confused the model.

3.7 Detection Results

The final output of this system is the measured room status, which is then stored and displayed. The monitoring results are displayed in a web dashboard, as shown in Figure 4, which displays the room status. The top figure displays a graph of the room's status history. The bottom section details the current status for each monitored room, for example, in room 605: Its current status is "Terisi" (marked in red), with the last detection on

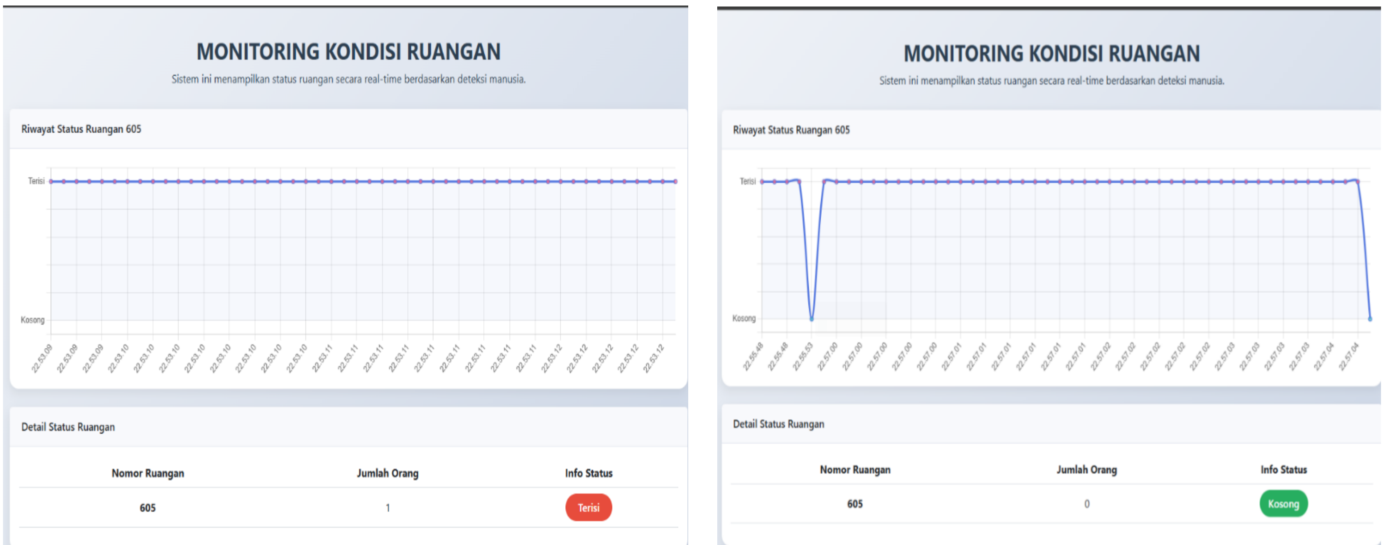

Figure 4. Monitoring Page

In (figure 5) shows how the display of this system works in detecting humans where the screen contains information about the room, the number of people and the status of the room. To calculate the number of people, the max() calculation logic is used by calculating how many bodies and faces are detected and then taking the largest number between the number of bodies and faces detected, in the countdown calculation it will count down from 60 to 0 to change the status which was initially if the "Ruang Kosong" was green will change if the "Ruangan Terisi" became red.

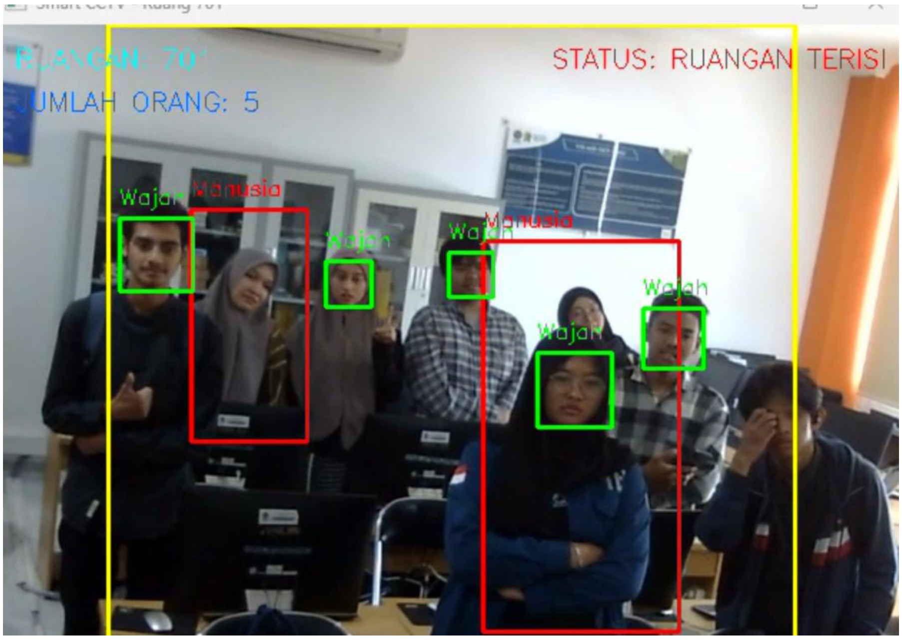

Figure 5. Screenshot of ROI System Video Window

4. CONCLUSSION

This study demonstrates that the implementation of a static Region of Interest (ROI) significantly improves computational efficiency in a human detection-based indoor monitoring system. An integrated architecture combining multiple detection, timing logic, and centralized data presentation was successfully implemented. The key findings are as follows:

1. The static ROI successfully reduced the computational load and minimized false positives originating outside the focus area. Detection state stability was also effectively improved through the implementation of minimum duration logic and a tolerance period.
2. The dual detection approach (body with HOG and face with Haar Cascade) combined with max() logic proved reliable in determining the number of people, with face detection demonstrating superiority in overlapping object scenarios.
3. A major weakness identified was the static ROI's inability to filter out visual clutter (e.g., shadows, reflections) within the monitoring zone. This resulted in a very low Precision value (0.04), confirming that this method functions as a spatial filter, rather than a contextual one.

REFERENCES

[1] D. Fuentes *et al.*, "IndoorCare: Low-Cost Elderly Activity Monitoring System through Image Processing," *Sensors*, vol. 21, no. 18, p. 6051, Sep. 2021, doi: 10.3390/s21186051.
[2] M. Z. Li, G. Liu, Z. Mao, Q. S. Yang, and J. W. Gu, "Two-dimensional motion estimation using phase-based image processing with Riesz transform," *Mech Syst Signal Process*, vol. 188, p. 110044, Apr. 2023, doi: 10.1016/j.ymssp.2022.110044.
[3] P. Wu *et al.*, "Next-generation machine vision systems incorporating two-dimensional materials: progress and perspectives," *InfoMat*, vol. 4, no. 1, p. e12275, 2022.
[4] H. P. Aguero, "Review of the Current Technologies and Applications of Digital Image Processing," *Journal of Biomedical and Sustainable Healthcare Applications*, vol. 2, no. 2, pp. 148–158, 2022.
[5] S. Sumijan and P. A. W. Purnama, "Teori dan Aplikasi Pengolahan Citra Digital Penerapan dalam Bidang Citra Medis," 2021, *PENERBIT INSAN CENDEKIA MANDIRI.*
[6] M. S. Hossain *et al.*, "Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images," *Sci Rep*, vol. 13, no. 1, p. 11314, Jul. 2023, doi: 10.1038/s41598-023-38109-6.
[7] Z. Tian, X. Si, Y. Zheng, Z. Chen, and X. Li, "Multi-step medical image segmentation based on reinforcement learning," *J Ambient Intell Humaniz Comput*, vol. 13, no. 11, pp. 5011–5022, 2022.
[8] D. Widiyanto, "Tinjauan Algoritma RoI (Region of Interest) Dengan Metode Pengambangan Otsu Dan Klasterisasi K-Mean; Hasil Dan Tantangannya," *Informatik: Jurnal Ilmu Komputer*, vol. 16, no. 2, pp. 75–86, 2020.
[9] C.-H. Choi, J. Kim, J. Hyun, Y. Kim, and B. Moon, "Face detection using haar cascade classifiers based on vertical component calibration," *Human-centric Computing and Information Sciences*, vol. 12, no. 11, pp. 1–17, 2022.

[10] M. Aishwarya and N. Neelima, “The analogy of HAAR cascade and HOG approaches for facial emotion recognition,” in *Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT*, Springer, 2021, pp. 699–707.

[11] M. Chandrakala and P. D. Devi, “Two-stage classifier for face recognition using HOG features,” *Mater Today Proc*, vol. 47, pp. 5771–5775, 2021.

[12] D. R. Rochmawati, “DETEKSI WAJAH DENGAN METODE HAAR CASCADE MENGGUNAKAN OPENCV (FACE DETECTION WITH HAAR CASCADE METHOD USING OPENCV),” *Jurnal Teknologi Komputer dan Informatika*, vol. 3, no. 1, 2024.

[13] F. T. Anggraeny, B. Rahmat, and S. P. Pratama, “Deteksi Ikan Dengan Menggunakan Algoritma Histogram of Oriented Gradients,” *Inform. Mulawarman J. Ilm. Ilmu Komput*, vol. 15, no. 2, p. 114, 2020.

[14] I. K. S. Buana, “Aplikasi untuk pengoprasian komputer dengan mendeteksi gerakan menggunakan opencv python,” 2018.

[15] M. Ahmed, M. D. Salman, R. Adel, Z. Alsharida, and M. Hammood, “An intelligent attendance system based on convolutional neural networks for real-time student face identifications,” *Journal of Engineering Science and Technology*, vol. 17, no. 5, pp. 3326–3341, 2022.

[16] N. M. Alruwais and M. Zakariah, “Student recognition and activity monitoring in e-classes using deep learning in higher education,” *IEEE access*, vol. 12, pp. 66110–66128, 2024.

[17] D. B. Hidayat, B. M. Nugraha, D. A. D. Bagaskara, D. P. Widyadhana, D. Purwitasari, and I. K. E. Purnama, “A Classroom Usage Monitoring System with Image Detection for Student Attendance,” in *2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT)*, IEEE, 2024, pp. 7–12.

[18] N. J. Wala’a and J. M. Rana, “A survey on segmentation techniques for image processing,” *Iraqi Journal for Electrical and Electronic Engineering*, vol. 17, no. 2, pp. 73–93, 2021.

[19] R. F. Falah, O. D. Nurhayati, and K. T. Martono, “Aplikasi pendeteksi kualitas daging menggunakan segmentasi region of interest berbasis mobile,” *Jurnal Teknologi dan Sistem Komputer*, vol. 4, no. 2, pp. 333–343, 2016.

[20] T. Long *et al.*, “A generic pixel pitch calibration method for fundus camera via automated ROI extraction,” *Sensors*, vol. 22, no. 21, p. 8565, 2022.