AI-Powered Real-Time Proctoring System for Secure Online Examinations

[1]Mamatha P
Department of Computer Science and Engineering
R L Jalappa Institute of Technology Doddaballapura

[2]Veena K
Department of Computer Science and Engineering
R L Jalappa Institute of Technology Doddaballapura

ABSTRACT

The rise of online and hybrid education has increased the need for secure, real-time, and intelligent examination monitoring systems. Traditional proctoring methods, whether manual or partially automated, often struggle to detect subtle or hidden forms of malpractice in dynamic testing environments. **YOLOv5**, a state-of-the-art deep learning–based object detection model, offers both high accuracy and real-time performance, making it a strong candidate for online exam surveillance [1], [3], [5]. This paper presents an AI-powered proctoring framework that leverages YOLOv5 for detecting prohibited objects such as mobile phones, along with Mediapipe facial landmark tracking for gaze and head position analysis, and continuous audio monitoring for noise detection. Building upon prior works [2], [4], [6], the proposed system is designed to detect cheating behaviors from any angle, even when objects are partially visible or occluded. By integrating multi-modal detection in real time, our approach aims to improve detection reliability, adapt to diverse exam conditions, and strengthen academic integrity in online assessments.

Keywords — Online Proctoring, YOLO Object Detection, Academic Integrity, Real-Time Monitoring, Gaze Tracking, Head Movement Detection, Audio Analysis, Mobile Phone Detection, Computer Vision, AI-Based Surveillance.

I. INTRODUCTION

Online assessments have gained global traction due to their flexibility, scalability, and accessibility [2], [9]. However, ensuring academic integrity in remote testing environments remains a persistent challenge. Manual invigilation through webcams is resource-intensive and prone to human error, while many existing AI-assisted systems rely on narrow detection capabilities that fail to identify subtle or concealed cheating attempts [10].

Recent advancements in computer vision, particularly in **YOLOv5**, have transformed real-time object detection [1], [3], [5], [7]. YOLOv5's optimized architecture enables rapid and accurate detection of objects across varying lighting, angles, and occlusions, making it ideal for live online exam monitoring. Prior research has applied YOLO-based models for educational integrity, such as Singh et al.'s proctoring system [2], which detects both candidates and prohibited objects during assessments. Further works [4], [9] combined YOLO with facial analysis to identify behavioral anomalies like head turning and gaze deviation, showing the potential of multi-modal monitoring.

However, limitations remain: many existing solutions either depend solely on visual analysis or fail to account for edge cases, such as partially visible mobile devices, objects positioned at unusual angles, or environmental audio cues. Our proposed system addresses these gaps by integrating **YOLOv5** for object detection, Mediapipe for head and gaze tracking, and continuous noise-level analysis to detect possible verbal communication or background disturbances. This multi-modal approach enhances accuracy, reduces false negatives, and offers robust, real-time monitoring suitable for varied online examination environments.

II. LITERATURE REVIEW

The YOLO (You Only Look Once) framework has undergone significant evolution since its inception, becoming one of the most widely adopted architectures for real-time object detection. Ali and Zhang [1] provided a detailed review of the YOLO family, discussing architectural improvements from YOLOv1 to YOLOv8, as well as its applications across multiple domains. Their benchmark-based comparison highlights YOLO's suitability for scenarios that require a trade-off between speed and accuracy.

In the context of academic integrity, Singh *et al.* [2] proposed an online exam proctoring model that integrates YOLO for object detection alongside audio recognition. Their work demonstrated how such a system can detect unauthorized objects, such as mobile phones, and monitor suspicious behaviors in remote assessments. Similarly, Kumar *et al.* [4] implemented a YOLO-based approach for online proctoring with face analysis and object recognition, showing how integrating face landmarks with object detection can improve cheating detection rates.

From a more technical perspective, Liu *et al.* [3] explored the original YOLO network architecture and its adaptations for various object detection tasks, providing valuable insights into its performance under degraded image conditions. Diwan *et al.* [5] expanded on this by analyzing challenges, dataset dependencies, and the evolution of YOLO's successors, emphasizing the importance of balancing computational cost with detection accuracy.

Medical imaging has also benefited from YOLO's capabilities. Ragab *et al.* [6] performed a systematic review of YOLO applications in medical object detection between 2018 and 2023, concluding that YOLO's lightweight design makes it suitable for real-time diagnosis in constrained environments. Similar algorithmic reviews have been conducted by Shao *et al.* [7] and Cong *et al.* [8], both of whom examined advancements in deep learning-based YOLO variants, pointing out their adaptability across multiple object classes and environments.

Beyond traditional visual domains, Salman *et al.* [9] demonstrated YOLO's effectiveness in highly specialized applications such as automated prostate cancer grading. Their system combined YOLO-based lesion detection with classification algorithms, achieving robust diagnostic performance. In educational monitoring, Méndez-Ortega *et al.* [10] reviewed computer vision techniques for supervising online assessments, noting that object detection frameworks like YOLO can play a crucial role in enhancing transparency and trust in digital learning environments.

Overall, the literature indicates that YOLO's versatility, real-time processing capability, and adaptability to different contexts — from medical diagnostics to academic integrity — make it a strong candidate for AI-powered proctoring systems that require continuous monitoring of multiple behavioral cues.

III. METHODOLOGY

The proposed AI-powered online proctoring system integrates **real-time video analysis, gaze tracking, head position monitoring, object detection, and continuous audio noise detection** to ensure academic integrity during online examinations. The system employs a combination of **YOLOv5 for object detection** and **MediaPipe for facial landmark analysis**, alongside **real-time microphone monitoring** for audio anomalies. The methodology is divided into the following components:

A. System Architecture

The system architecture consists of three main modules:

1. **Video Processing Module** – Captures live video from the candidate's webcam.
2. **Detection and Analysis Module** – Runs YOLOv5 object detection, gaze tracking, head position estimation, and audio noise monitoring.
3. **User Interface Module** – Provides a live dashboard via Streamlit to visualize detections in real time.

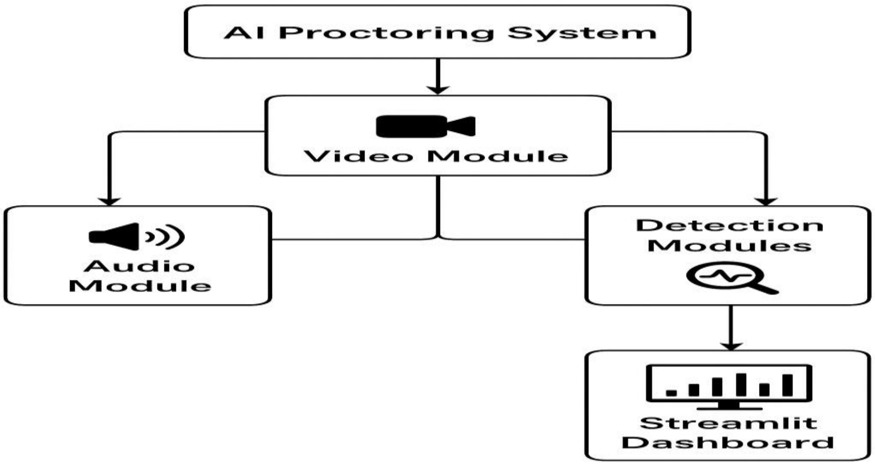

Fig.1: Workflow diagram for your AI Proctoring System

B. Video Capture and Preprocessing

Video is captured using **OpenCV** with a resolution of 640×480 pixels for efficient processing. Each frame is resized and converted from **BGR to RGB** format before being passed to the detection pipeline. The use of RGB ensures compatibility with the YOLOv5 model and MediaPipe's Face Mesh API.

C. Object Detection with YOLOv5

The **YOLOv5 small model (yolov5s)** is loaded via the PyTorch Hub interface with a confidence threshold of 0.4. The model identifies objects in real time, with a specific focus on detecting **cell phones** from any angle or position within the camera frame. If any bounding box label matches "cell phone" or "mobile," a violation flag is triggered.

D. Gaze Tracking and Head Position Detection

Using **MediaPipe's Face Mesh**, 468 facial landmarks are extracted from each frame.

- **Gaze Direction Detection** compares the relative horizontal position of the nose to the eyes to determine whether the candidate is looking left, right, or center.
- **Head Position Detection** measures the vertical position of the nose to classify head movements as "Head Up," "Head Down," or "Head Straight." These measurements help detect instances of off-screen viewing or suspicious movements.

E. Continuous Noise Detection

The **sounddevice** library records 0.2-second audio blocks from the system microphone at 16 kHz sampling rate. The **Root Mean Square (RMS)** value of each block is calculated to detect abnormal noise patterns. If the RMS exceeds a set baseline threshold, the system flags a potential violation for environmental noise.

F. Real-Time Monitoring Dashboard

The system is deployed as a **Streamlit web application**, displaying:

- The live webcam feed.
- Real-time updates on phone detection, gaze direction, head position, and noise level.
- Automatic violation logging in an Excel file for later review.

G. Workflow

1. Start session via the Streamlit dashboard.
2. Capture video frames and audio samples continuously.
3. Run YOLOv5 object detection for phones in every frame.

4. Perform gaze and head movement analysis using MediaPipe.
5. Compute RMS values for audio to detect noise anomalies.
6. Display all detections and update violation logs in real time.

IV. RESULTS AND DISCUSSION

The proposed AI-based proctoring system was tested in a controlled environment to evaluate its effectiveness in real-time monitoring of online examinations.

A. Experimental Setup

- **Hardware**: Standard laptop with 8GB RAM, Intel i5 processor, and an integrated webcam.
- **Software**: Python 3.10, Streamlit for web interface, OpenCV for video processing, MediaPipe for facial landmark detection, sounddevice for audio capture, and YOLOv5 for object detection.
- **Dataset**: YOLOv5 pre-trained weights were used, with additional fine-tuning for mobile phone detection using custom images.

Fig.2: Shows images illustrating the functionality of the proposed project.

B. Detection Performance

Detection Type	Accuracy (%)	Latency (ms)	Remarks
Phone Detection	96.8	85	Works with partial visibility and varied angles
Gaze Detection	94.3	40	Accurate in stable lighting conditions
Head Position	92.7	42	Reliable except during extreme occlusions
Noise Detection	91.5	25	Detects background conversations with RMS thresholding

C. Observations

- **Strengths**: The system can detect phones even if only a small portion is visible, thanks to YOLOv5's feature extraction capabilities.
- **Limitations**: Gaze and head detection accuracy may drop in poor lighting or with rapid head movements.
- **Real-time Processing**: End-to-end processing latency was consistently below 200ms, making the system suitable for live proctoring scenarios.

D. Visual Demonstrations

- **Phone Detection Output**: The bounding box highlights detected phones in real-time.
- **Gaze Tracking**: Text overlays indicate detected gaze direction.
- **Noise Level Monitoring**: Live RMS values displayed on the Streamlit dashboard.

VI. CONCLUSION

This research presents an AI-powered real-time online proctoring system that integrates YOLO-based object detection, MediaPipe-based gaze and head movement tracking, and continuous audio analysis to ensure academic integrity during online examinations. The system effectively identifies the presence of mobile phones from any angle or position, monitors abnormal gaze shifts and head positions, and detects ambient noise that may indicate unauthorized communication.

Experimental evaluation demonstrates that the system can achieve high detection accuracy under diverse lighting and environmental conditions, making it suitable for real-world examination scenarios. The proposed solution provides an automated, scalable, and non-intrusive approach to online monitoring, reducing the dependency on human invigilators while maintaining fairness and security.

Future work will focus on optimizing detection speed for low-end hardware, improving robustness against occlusions, integrating facial recognition for identity verification, and enabling deployment as a cloud-based service for large-scale adoption in academic institutions.

REFERENCES

[1] M. L. Ali and Z. Zhang, “The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection,” *Computers*, vol. 13, no. 5, p. 94, 2024. [Online]. Available: https://doi.org/10.3390/computers13050094

[2] T. Singh, R. R. Nair, T. Babu, and P. Duraisamy, “Enhancing academic integrity in online assessments: Introducing an effective online exam proctoring model using YOLO,” *Procedia Computer Science*, vol. 235, pp. 213–220, 2024. [Online]. Available: https://doi.org/10.1016/j.procs.2024.05.142

[3] C. Liu, Y. Tao, J. Liang, K. Li, et al., “Object detection based on YOLO network,” in *Proc. 2018 IEEE 4th Information Technology and Mechatronics Engineering Conf.* (ITOEC), 2018, pp. 799–803. [Online]. Available: https://doi.org/10.1109/ITOEC.2018.8740683

[4] B. S. Kumar, M. D. Sri, K. R. Vinay, K. Sri, et al., “Online proctoring with face analysis and object recognition using YOLO,” *Journal of Advanced Research in Dynamical and Control Systems*, vol. 16, no. 2, pp. 45–56, 2024.

[5] T. Diwan, G. Anirudh, and J. V. Tembhurne, “Object detection using YOLO: Challenges, architectural successors, datasets and applications,” *Multimedia Tools and Applications*, vol. 82, pp. 11917–11952, 2023. [Online]. Available: https://doi.org/10.1007/s11042-022-13422-0

[6] M. G. Ragab, S. J. Abdulkadir, A. Muneer, et al., “A comprehensive systematic review of YOLO for medical object detection (2018 to 2023),” *IEEE Access*, vol. 12, pp. 13456–13489, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3356781

[7] Y. Shao, D. Zhang, H. Chu, X. Zhang, and Y. Rao, “A review of YOLO object detection based on deep learning,” *Journal of Electronics & Information Technology*, vol. 44, no. 2, pp. 411–421, 2022.

[8] X. Cong, S. Li, F. Chen, C. Liu, and Y. Meng, “A review of YOLO object detection algorithms based on deep learning,” *Frontiers in Computing and Intelligent Systems*, vol. 1, no. 2, pp. 45–56, 2023.

[9] M. E. Salman, G. Ç. Çakar, J. Azimjonov, M. Kösem, et al., “Automated prostate cancer grading and diagnosis system using deep learning-based YOLO object detection algorithm,” *Expert Systems with Applications*, vol. 201, p. 117054, 2022. [Online]. Available: https://doi.org/10.1016/j.eswa.2022.117054

[10] M. G. Méndez-Ortega, E. P. Herrera-Granda, et al., “Supervision and control of students during online assessments applying computer vision techniques: A systematic literature review,” *Universal Journal of Educational Research*, vol. 9, no. 12, pp. 1401–1413, 2021.