International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Online Proctoring System: A Review in Deep Learning – Based Online Proctoring System Using Object Recognition, People Counting, Face Spoofing, Face Recognition & Detection and Head & Eye Tracking.

Abel Thomas[1], Ayisha A A[2], George Thomas [3], Meenakshy R[4], Dini Davis[5]

[1,2,3,4] Student, Department of Computer Science and Engineering, IES College of Engineering, Chittilappilly, Thrissur.
[5]Assistant Professor, CSE Department, IES College of Engineering, Chittilappilly, Thrisssur

ABSTRACT

In this paper we dedicated to developing a review on Online Proctoring System aimed at enabling remote invigilation with a focus on significantly enhancing the efficacy of preventing malpractices during online examinations. Traditionally, many institutions have implemented manual online proctoring methods, but these approaches have shown limited success in deterring academic misconduct. In response, this paper prioritizes a comprehensive set of features essential for effective proctoring, mirroring the functionalities available during traditional offline examinations.

Key features incorporated into our project include eye and head tracking, face recognition, face detection, people counting, object recognition, and face spoofing detection. By leveraging advanced technologies, system is designed to detect various forms of movement , usage of gadgets (like mobile phones , smart watch) , checking the number of peoples and eye movement, providing invigilators with real-time insights.

In case of any misconduct, the system issues timely and reasonable warnings to the students, promoting a fair and secure examination environment. The primary objective of this paper is to empower invigilators to proctor remotely while addressing the shortcomings of conventional online proctoring methods. Our team is committed to delivering a robust and user friendly solution that not only detects and prevents malpractices effectively but also ensures a seamless and trustworthy examination experience for both students and educators.

Keywords: Online proctoring system, Education, Authentication, Abnormal behavior detection; face detection; face recognition; eye and head tracking; object detection; face spoofing.

1.INTRODUCTION

In this review paper we aim to revolutionize the landscape of remote examinations by leveraging modern technologies for secure and efficient monitoring of online assessments. With the increasing prevalence of online learning, maintaining the integrity of the evaluation process is crucial. In this paper we discuss about the integrates advanced features such as Facial recognition, Facial detection, Head and Eye tracking, People counting, Face spoofing, Object detection to monitor examinee in real-time. By employing cutting-edge technologies, it ensures the identification of individuals and detects any irregularities or potential cheating during the

exam. This review paper portrays a comprehensive and innovative approach to addressing the challenges associated with remote examination supervision. In a rapidly evolving educational landscape, where online assessments have become common place , ensuring the integrity of exams is of paramount importance. In this paper leverages advanced technologies, including dlib for robust facial recognition and authentication, and YOLOv3 for efficient object detection, to create a sophisticated monitoring system. The integration of dlib enhances the system's ability to accurately identify and verify test-takers, contributing to a secure authentication process. YOLOv3, with its real-time object detection capabilities, enables the system to track and analyze various elements within the examination environment, identifying any suspicious activities or potential instances of academic dishonesty. The holistic combination of these technologies ensures a reliable and adaptable online proctoring solution, capable of addressing a diverse range of testing scenarios. By enhancing the security and credibility of remote assessments, it provide educational institutions and organizations with a powerful tool to maintain the integrity of online exams and foster a trustworthy evaluation process in the digital era.

2. LITERATURE REVIEW

In this paper, we intend to discuss about a web based automated examination system which can detect any malicious activities and flag them, to ensure a fair proposition of exams. Essentially, we detect malpractices by incorporating computer audio and visual movements using webcam and microphone.

While it allows students to take a test from a college lab with specific technical prerequisites, it also removes the need for physical examination centers. Primarily, Vision based tracking consists of Eye ball tracking, Lip movement, Face spoofing, Mobile phone detection, Additional member detection in frame and more.

S.Prathish et al. [1] used the model-based head pose estimation method and the audio-based detection method to complete the test abnormal behavior detection. However, the accuracy rate of the head pose estimation of this method is nothigh enough, and the use of a microphone to collect sound can infringe the relevant privacy of examinees. Moreover, the abnormal behavior detection process does not consider eye tracking and mouth movement analysis.

In [2], the authors proposed a multimedia analytics system for online exam proctoring. With the captured videos and audio, they extract low-level features from six basic components: text detection, user verification, active window detection, speech detection, gaze estimation, and phone detection. These features are then processed in a temporal window to acquire highlevel features and then used for cheat detection. However, the system is not feasible as it requires the examinee to have a wearcam. Moreover, the solution does not have a face spoofing feature for user authentication.

Hu et al. [3] proposed a system that uses an image-based head pose estimation model and mouth movement analysis to discriminate the abnormal behavior of the examinee during the online examination. However, the system does not take eye-tracking functionality into consideration for analyzing the abnormal behavior of the examinee.

One of the common approaches to estimate head pose is using landmark points [4, 5, 6]. In this approach along with detected 2D landmarks, an average 3D mean mask and the intrinsic camera parameters are required to calculate the 3D head pose angle. Using this information, the extrinsic parameter of the camera is calculated which contains the information about 3D rotation and translation of face from the center of the camera. This approach has a few drawbacks. Firstly, the accuracy of this method heavily depends on the accuracy of the landmark model. The landmark model usually fails for large head pose angles since half of the features in the face are invisible. In such scenarios, the accuracy of head pose estimation will also drop. In addition to that, a mean 3D mask is used to perform the 3D to 2D alignment, and this will also introduce errors in head pose calculation. Since its accuracy drops when the examinee gives a large head pose, this approach is not suitable for our online proctoring system. Another common approach is using deep learning-based classification methods.

Some of the state-of-the-art models in this approach are Hopenet [7] and WHENet [8]. Hopenet uses Resnet50 as its backbone feature extractor, followed by a classifier that classifies each head pose angle. The network is trained using a combination of both classification and regression loss functions. WHENet also follows a similar approach but uses EfficientNet as a backbone feature extractor. Even though the two models mentioned here give a very accurate performance, they are not usable in our online proctoring system because of their computational complexity.

We require a model that can achieve real-time performance with sufficient accuracy to identify when an examinee is looking away from the screen. For training and validation of the head pose module, we used the Pandora dataset [9]. The dataset contains 100 annotated sequences collected from 10 male and 10 female subjects. Each subject has been recorded 5 times. For each subject, two sequences are performed with constrained movements, changing the yaw, pitch, and roll angles separately. Three additional sequences are completely unconstrained. The overall size of the dataset was around 130k.

For testing the head pose module, we used the BIWI [10] benchmark dataset. It contains 15k frames, with RGB (640×480) and depth maps (640×480). 20 subjects have been involved in the recordings: 4 of them were recorded twice, for a total of 24 sequences. The ground truth of yaw, pitch, and roll angles is reported together with the head center and the calibration matrix.

These studies collectively contribute to the evolving landscape of surgical lighting systems. From historical perspectives to cutting-edge technologies, the literature reviewed highlights the diverse approaches aimed at improving illumination, efficiency, and precision in surgical settings. The integration of automation, advanced sensors, and innovative designs underscores the ongoing efforts to redefine and enhance the surgical experience.

3. RESEARCH TOPIC

3.1 People Counting

For counting the number of people, the robust YOLOv3 object detector from OpenCV takes center stage, effectively serving as the cornerstone for detecting and quantifying the presence of individuals within a given frame. Specifically, the primary objective is to employ this cutting- edge technology to identify and count the number of people present in the visual field under examination. The YOLOv3 model excels in this task by simultaneously predicting bounding boxes and class probabilities, allowing for efficient and accurate detection of multiple objects, including individuals, within a single pass. The cheating detection mechanism operates by analyzing the continuous stream of frames. If, over a span of 10 consecutive frames, the YOLOv3 detector fails to identify any individuals or detects more than one person, it triggers an alert indicating a potential instance of academic dishonesty. This stringent condition is imposed to ensure a high level of confidence in the accuracy of the person detection process and, consequently, the reliability of the cheating detection mechanism.

3.2 Face Detection

In this paper for face detection, we harnessed the power of OpenCV's DNN (Deep Neural Network) module to implement a sophisticated facial detection mechanism. This crucial component is instrumental in identifying and locating the examinee's face within the examination frame. The utilization of DNN technology ensures the integration of deep learning principles, enhancing the accuracy and efficiency of the face detection process. The underlying face detection model is constructed upon the Single Shot Detector (SSD) framework, a pioneering approach in object detection within images. SSD excels in real-time applications by efficiently predicting bounding boxes and class probabilities for multiple objects in a single pass. This framework is particularly well-suited for our online proctoring system, where rapid and accurate identification of the examinee's face is paramount.

3.3 Face Recognition

For the identification of examinees, we integrates Dlib's face verification model, leveraging its pre-trained ResNet50 Convolutional Neural Network (CNN) model. This idea of advanced model is employed to extract a highly discriminative 128 dimensional feature vector from facial images stored in the database. Upon detecting the examinee's face in real-time, the same process is repeated to generate a 128D feature vector specific to the detected face. The critical step in this procedure involves calculating the Euclidean distance between these two feature vectors. If the computed distance falls below a predetermined threshold, in our case, Dlib's default threshold of 0.6, it signifies a high degree of similarity, and the faces are deemed to match. This robust face verification mechanism ensures the accurate association of an examinee with their respective identity, contributing to the overall security and reliability of the online proctoring system.

3.4 Head Pose Estimation

One of the common approaches to estimate head pose lightweight head pose estimation model tailored for real-time performance in systems with limited computational resources. Our approach enables the accurate prediction of pitch, yaw, and roll angles directly from a cropped face, eliminating the need for landmarks or depth maps. Opting to train a pose estimation network from scratch rather than relying on landmarks, we identified significant advantages associated with deep networks. Unlike landmark-based methods, deep networks are not contingent on a specific head model, landmark detection approach, or subset of points for alignment. Additionally, they consistently produce pose predictions, mitigating issues encountered by landmark-based methods when detection fails, particularly in extreme poses. This decision enhances the adaptability and reliability of our pose estimation model within resource-constrained environments.

3.5 Eye Tracking

We used Dlib's pre-trained network for detecting and predicting 68 facial landmarks on the examinee's face. Left eye is defined by the following landmarks - 36,37,38,39,40,41. Right eye is defined by the following landmarks - 42,43,44,45,46,47. First, we segmented the eye regions by using a mask. Then we applied binary thresholding on the eye regions to separate the eyeballs from the rest of the eye regions. Eyeballs become black and the rest regions stay white. Then a vertical separator was created at the middle of each eye.Finally, to determine if the examinee is looking left or right, we defined an eye-tracking ratio as:

AvgET R = RightEyeET R + Lef tEyeET R 2 where,

RightEyeET R = No. of white pixels on lef t side\

No. of white pixels on right side Lef tEyeET R = No. of white pixels on lef t side \

No. of white pixels on right side

After extensive trial and testing, we fixed the following thresholds for the AvgETR: ⩽0.35 (looking outside the screen), 0.36 to 3.9 for the center (looking at the screen), ⩾4 for left (looking outside screen). If the examinee is looking outside the screen for more than 10 consecutive frames, then the examinee is said to be cheating

3.6 Object Detection

We integrated OpenCV's YOLOv3 object detector for an additional layer of security, specifically to identify instances of prohibited items such as mobile phones, laptops, TVs, and books. YOLOv3's advanced object detection capabilities allow it to efficiently scan the examination environment, pinpointing and categorizing these banned items in real-time. The system is designed to trigger a cheating alert if one or more instances of a forbidden item are detected persistently for more than 10 consecutive frames. This stringent condition ensures that the system responds to sustained and potentially dishonest behavior, providing a robust mechanism to uphold the integrity of the examination process. By leveraging YOLOv3 for item detection, our online proctoring system not only addresses the issue of unauthorized materials but also adds a preventive layer against potential cheating scenarios.

3.7 Face Spoofing

To identify whether the examinee is real or a photograph, we implemented face spoofing functionality. After capturing the examinee's face image using the Face Detection module, it is further converted into YCrCb and CIE L*u*v* color spaces using OpenCV. Later, histograms are calculated from both the color spaces and concatenated together. The concatenated histogram is sent to Scikit-learn's ExtraTreesClassifier model for classifying face into real/spoof. If the face is classified as a spoof for more than 10 consecutive frames, then the examinee is said to be cheating.

4. ADVANTAGES OF ONLINE PROCTORING SYSTEM

The Online proctoring offers several advantages for conducting exams, addressing various challenges associated with traditional testing methods. One of its primary benefits is enhanced accessibility, allowing exams to be taken remotely and overcoming geographical barriers that may hinder participation. This fosters inclusivity and expands the reach of educational assessments. Moreover, online proctoring introduces a level of flexibility that accommodates diverse time zones and student schedules. This adaptability in scheduling exams caters to the needs of a global and diverse student population, promoting a more inclusive learning environment. A crucial aspect of online proctoring is real-time monitoring, which serves as a deterrent to cheating during exams. Features such as facial recognition and screen monitoring contribute to heightened exam security, ensuring the integrity of the assessment process. The transition to online proctoring also brings about cost-efficiency by reducing the reliance on physical testing centers. This not only saves costs associated with traditional proctoring methods but also streamlines the examination process for both institutions and students. Convenience is another notable advantage, as online proctoring eliminates the need for travel. This makes the examination process more convenient for both students and institutions, further enhancing the overall efficiency of the assessment process. The integration of data analytics is a

valuable feature, providing institutions with insights for performance assessment and continuous improvement of the online proctoring process. This data-driven approach contributes to refining examination procedures and ensuring their effectiveness. Customization options empower institutions to tailor proctoring settings based on specific exam requirements and security needs. This adaptability ensures that the online proctoring solution aligns with the unique demands of each educational context. Additionally, online proctoring reduces the administrative burden by automating the monitoring process. This automation diminishes the need for manual oversight, saving valuable administrative time and resources.

The inclusion of audio monitoring is a proactive measure to detect and discourage unauthorized communication during exams, adding an extra layer of security to the assessment process. Real- time intervention capabilities enable proctors to respond promptly to any suspicious behavior detected during exams. This real-time responsiveness helps maintain the integrity of the exam and ensures a fair and secure testing environment. The incorporation of innovative technologies such as artificial intelligence (AI) and machine learning enhances detection capabilities. This proactive approach helps stay ahead of evolving cheating methods, ensuring the continuous improvement and effectiveness of online proctoring solutions.

5.RESULTS

The images are trained from the Pandora dataset. The model takes a face crop rescaled to 100 x 100 as input. The face detection bounding box output is enlarged by 100% and the resulting bounding box is used to crop out thehead region and pass that as input to the network. We used the Adam optimizer with a learning rate of 0.001 to train the model. The model was either trained for 100 epochs or was used with early stopping which monitored the validation loss with the patience of 20 epochs. The corresponding training loss and validation loss achieved for the best model. The model was trained using Mean Squared Error (MSE) as a loss function and evaluated on the test dataset using Mean Average Error (MAE). To prevent the model from overfitting we used dropout regularization after each convolution and dense layer.

6.CONCLUSION

In conclusion, the implementation of an online proctoring system brings forth enhanced security and integrity to the examination process. Through real-time monitoring, facial recognition, and advanced AI algorithms, the system ensures a fair and trustworthy environment for online assessments. The online proctoring system stands as a pivotal tool in maintaining academic integrity and fostering a secure digital learning experience. From the study we came to the conclusion that in the existing system, the accuracy rate of head pose estimation is not high enough, and the use of the microphone to collect sound can infringe there levant privacy examinees. To compensate the drawbacks of the existing system we are implementing highly accurate head pose estimation and also avoiding microphone to avoid privacy issues of the examinee.

7. REFERENCES

[1]: Swathi Prathish, Kamal Bijlani, et al., “An intelligent system for online exam monitoring,” in 2016 International Conference on Information Science (ICIS). IEEE,2016, pp. 138–143

[2]: Yousef Atoum, Liping Chen, Alex X Liu, Stephen DH Hsu, and Xiaoming Liu, “Automated online exam proctoring,” IEEE Transactions on Multimedia, vol. 19, no.7, pp. 1609–1624, 2017. [3] : Senbo Hu, Xiao Jia, and Yingliang Fu, “Research on abnormal behavior detection of online examination based on image information,” in 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, 2018, vol. 2, pp. 88–91.

[4]: Vahid Kazemi and Josephine Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 18671874.

[5]: Adrian Bulat and Georgios Tzimiropoulos, “How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks),” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1021–1030.

[6]: Xiangyu Zhu, Xiaoming Liu, Zhen Lei, and Stan Z. Li, “Face alignment in full pose range: A 3d total solution,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 78 92, Jan 2019.

[7] : Nataniel Ruiz, Eunji Chong, and James M. Rehg, “Finegrained head pose estimation without keypoints,”2018.

[8] : Yijun Zhou and James Gregson, “Whenet: Real-time fine-grained estimation for wide range head pose,” 2020.

[9] : Guido Borghi, Matteo Fabbri, Roberto Vezzani, Simone Calderara, and Rita Cucchiara, “Facefrom-depth for head pose estimation on depth images,” IEEE transactions on pattern analysis and machine intelligence, vol.42, no. 3, pp. 596–609, 2018.

[10]: Gabriele Fanelli, Matthias Dantone, Juergen Gall, Andrea Fossati, and Luc Van Gool, “Random forests for real time 3d face analysis,” International journal of computer vision, vol. 101, no. 3, pp. 437– 458, 2013.