IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Test Guardian Using Computer Vision

[1]Aditya Bothe, [2]Raghav Patil, [3]Vaibhav Raktate, [4]Barkha Kumari, [5]Suyash Gadhave

[1]Developer, [2]Developer, [3]Developer, [4]Professor, [5]Developer

[1]Computer Science and Engineering (Artificial Intelligence and Machine Learning),

[1]G H Raisoni College of Engineering and Management, Pune, India

Abstract: The Covid-19 widespread has been one of the defining events in later history. It has influenced millions of lives and has had an effect on each division of civilization. No matter the space, the widespread has constrained it to actualize radical and imaginative changes. Instruction and the scholarly community have been identified as one such division that has been affected most adversely due to the widespread. Disturbing the age-old classroom setup, the widespread has constrained instructive institutions like schools and colleges to execute 'online classes'. However, the assessment angle of instruction remains to be wanted. Numerous programmed online exam proctoring frameworks have been proposed for online examinations amid this Covid-19 pandemic but they have certain impediments like less and wrong functionalities. In this paper, we construct a keen exam monitoring framework, which addresses numerous of the problems with past frameworks, to offer assistance teach maintain a strategic distance from malpractices during the exams.

Index Terms - Online proctoring system, Education, Authentication, Abnormal behavior detection

I. INTRODUCTION

Exams are a basic component of any instructive system. The impact of COVID-19 on instruction has caused many schools and colleges to switch their mode of exams from in-person exams to the online mode to follow to open safety regulations. Instructive Testing Benefit (ETS), the nonprofit instructive organization which offers standardized tests counting GRE and GMAT, is moreover permitting examinees to grant exams from domestic where they will be observed by a proctor for the entire length of the exam. Be that as it may, after multiple exams being conducted online, it is watched that examinees have scored astoundingly tall in tests due to a lack of the number of proctors. Many programmed online exam proctoring (OEP) systems have been proposed to handle this issue. But they mainly focus on the verification of examinees by performing tasks such as unique finger impression verification, confront confirmation, and voice recognition. However, the over strategies disregard the irregular behavior amid the examinations. Ordinary examinees confront the show screen to reply the questions. The unusual examination behavior ordinarily shows in the irregular changes in the head pose and eye development and the ceaseless opening and closing of the mouth. In this paper, we point to construct a programmed online exam proctoring framework that gives progressed verification and abnormal behavior observing of examinees in the online examination based on picture data. The input to our framework is a real-time video stream. The video stream is analyzed frame-by-frame and alerts are produced whenever pre-defined triggers are experienced. We make strides the confirmation prepare by counting a confront spoofing highlight. Through the utilize of head posture estimation, eye following, mouth development investigation, and the combination of certain choice rules, the observing of anomalous behavior such as turning heads and eyes and talking amid the online examination is completed.

II. LITERATURE REVIEW

2.1POPULATION AND SAMPLE

S.Prathish et al. [1] utilized the model-based head posture estimation strategy and the audio-based location strategy to total the test unusual behavior discovery. In any case, the precision rate of the head posture estimation of this strategy is not high sufficient, and the utilize of a receiver to collect sound can encroach the pertinent security of examinees. Besides, the abnormal behavior discovery handle does not consider eye tracking and mouth development analysis. In [2], the creators proposed a mixed media analytics framework for online exam proctoring. With the captured recordings and audio, they extricate low-level highlights from six essential components: content location, client confirmation, dynamic window discovery, discourse discovery, look estimation, and phone detection. These highlights are at that point handled in a worldly window to acquire high-level highlights and at that point utilized for deceive detection. However, the framework is not doable as it requires the examinee to have a wearcam. Besides, the arrangement does not have a confront spoofing highlight for client authentication. Hu et al. [3] proposed a framework that employments an image-based head posture estimation demonstrate and mouth development investigation to discriminate the unusual behavior of the examinee during the online examination. In any case, the framework does not take eye-tracking usefulness into thought for analyzing the abnormal behavior of the examinee.

2.2 Head Pose Estimation

One of the common approaches to evaluate head pose is utilizing point of intrigued centers [4, 5, 6]. In this approach along with detected 2D focuses of intrigued, an typical 3D pitiless cover and the intrinsic camera parameters are required to calculate the 3D head pose point. Utilizing this information, the outward parameter of the camera is calculated which contains the information around 3D insurgency and elucidation of stand up to from the center of the camera. This approach has a few drawbacks. Firstly, the accuracy of this procedure heightening depends on the accuracy of the point of intrigued appear. The point of intrigued illustrate as a run the show comes up brief for large head pose focuses since half of the highlights in the face are intangible. In such scenarios, the precision of head pose estimation will besides drop. In extension to that, a brutal 3D mask is utilized to perform the 3D to 2D course of action, and this will also introduce botches in head pose calculation. Since its accuracy drops when the examinee gives a tremendous head pose, this approach is not sensible for our online proctoring system. Another common approach is utilizing significant learning-based classification techniques. A few of the state-of-the-art models in this approach are Hopenet [7] and WHENet [8]. Hopenet uses Resnet50 as its spine incorporate extractor, taken after by a classifier that classifies each head pose point. The orchestrate is trained utilizing a combination of both classification and backslide hardship capacities. WHENet as well takes after a comparative approach but livelihoods EfficientNet as a spine incorporate extractor. Even though the two models said here convey a uncommonly accurate performance, they are not usable in our online proctoring system since of their computational complexity. We require a model that can finish real-time execution with sufficient accuracy to recognize when an examinee is looking missing from the screen.

Table 1. Subset of functionalities in our system

Category	Functionality
Authentication	Face Verification Face Spoofing
Abnormal behaviour detection	Image-based Head Pose Estimation Eye Tracking Mouth Movement Analysis

As per our knowledge, there is no previous work that has integrated all Authentication and Abnormal Behaviour Detection features provided in Table 1.

III. DATASET

For preparing and approval of the head posture module, we utilized the Pandora dataset [9]. The dataset contains 100 explained arrangements collected from 10 male and 10 female subjects. Each subject has been recorded 5 times. For each subject, two arrangements are performed with compelled developments, changing the yaw, pitch, and roll points independently. Three extra arrangements are totally unconstrained. The generally estimate of the dataset was around 130k. Illustration pictures from the Pandora dataset are appeared in Fig. 1

Fig. 1. Example images from Pandora dataset.

For testing the head posture module, we utilized the BIWI [10] benchmark dataset. It contains 15k outlines, with RGB (640 × 480) and profundity maps (640 × 480). 20 subjects have been included in the recordings: 4 of them were recorded twice, for a add up to of 24 arrangements. The ground truth of yaw, pitch, and roll points is detailed together with the head center and the calibration network. Illustration pictures from the BIWI dataset are appeared in Fig. 2

Fig. 2. Example images from BIWI dataset

To move forward the differences of the dataset and anticipate the show from overfitting we performed a arrangement of color and geometrical expansion procedures like even flip, irregular zooming, including arbitrary commotion (Gaussian, salt, and pepper), color jitter (arbitrary brightness varieties, irregular differentiate varieties, arbitrary immersion varieties), and irregular picture interpretations.

IV. PROPOSED SYSTEM

4.1Population and Sample

This work proposes a system that uses a webcam to monitor examinees during the online examination. The system architecture is shown in Fig. 3. After the camera captures the frame, banned item detection and person detection are performed. If one person is detected in the frame, then face detection is performed. The detected face is input to face spoofing, face verification, face landmark detection, and head pose estimation models. The detected landmarks from the landmark detection model are input to mouth movement analysis and eye-tracking. We analyze the output from all models to conclude whether the examinee is cheating or not. Table 2 contains all the functionalities present in our proposed system.

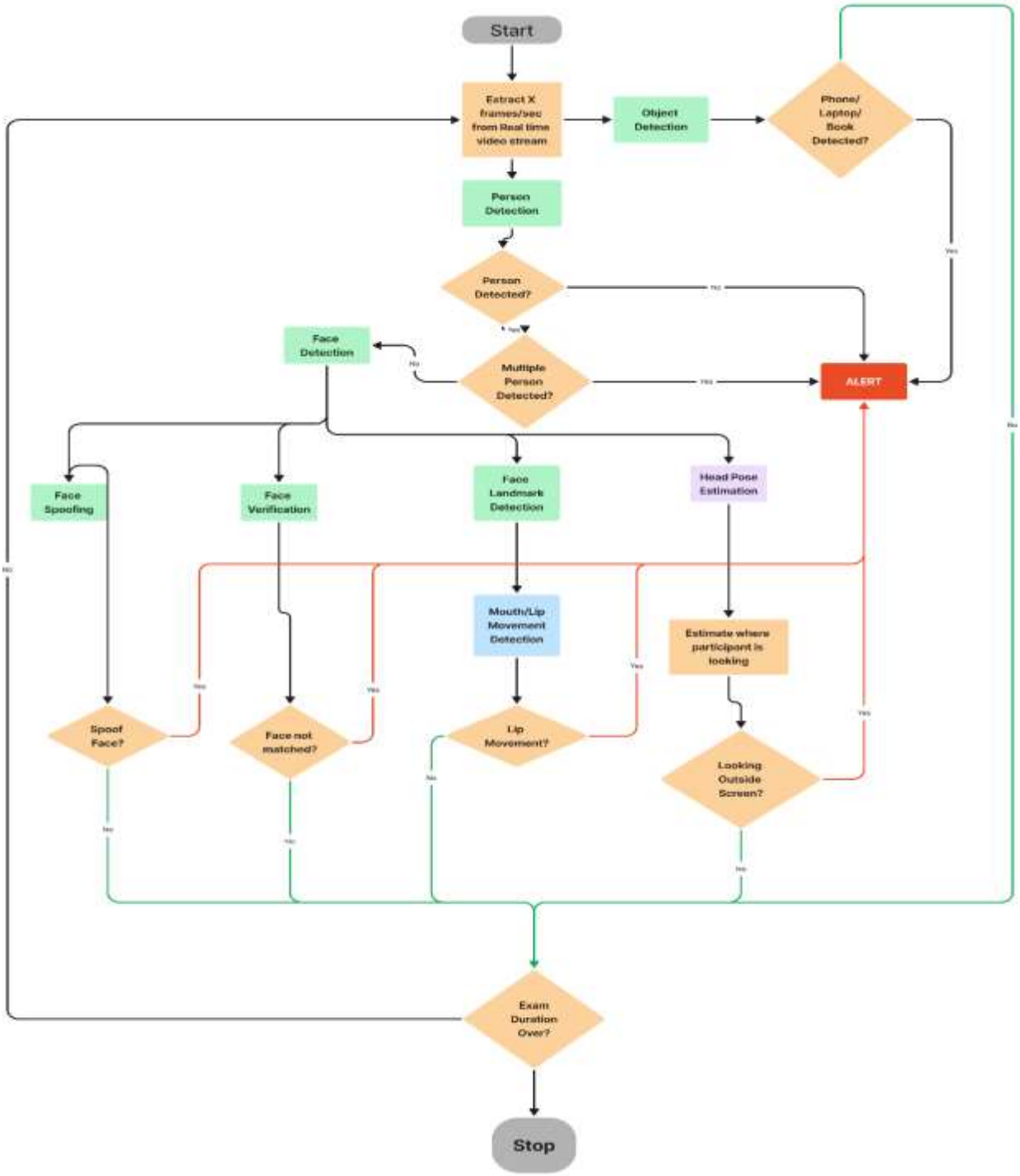

Fig. 3. Architecture of the proposed Online Proctoring System.

Table 2. Functionalities in our system

Category	Functionality	Technique
Base	Person Detection and Counting	Pre-trained
	Object Detection	Pre-trained
	Face Detection	Pre-trained
Authentication	Face Verification	Pre-trained
	Face Spoofing	Pre-trained
Abnormal Behavior Detection	Image-based Head Pose Estimation	Trained from Scratch
	Eye Tracking	Image-Processing
	Mouth Movement Analysis	Image-Processing

4.2Person Detection and Counting

We used OpenCV's [11] YOLOv3 object detector for detecting and counting the number of people in the frame. If no one or more than one person is detected for more than 10 consecutive frames, then the examinee is said to be cheating. Outputs from the Person Detection and Counting module are visible in Fig. 4. and Fig. 5.

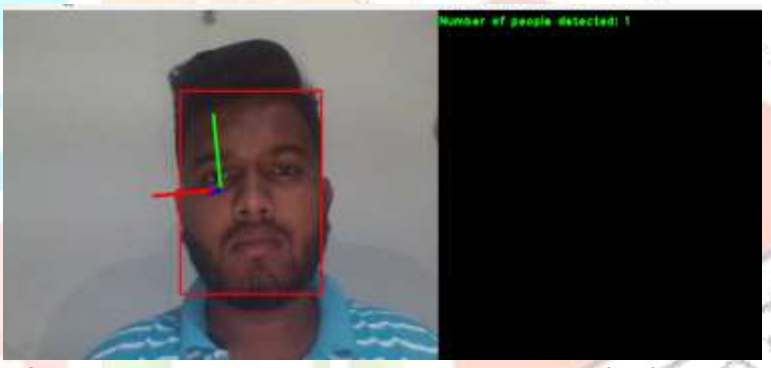

Fig. 4. Output from Person Detection and Counting module for frame with single person.

Fig. 5. Output from Person Detection and Counting module for frame with multiple people.

4.3 Object Detection

OpenCV's YOLOv3 object detector was also used for finding any instances of banned items including mobile phones, laptops, TV, and books. If one or more than one instance of any banned item is detected for more than 10 consecutive frames, then the examinee is said to be cheating. Outputs from the Object Detection module are shown in Fig. 6. and Fig. 7.

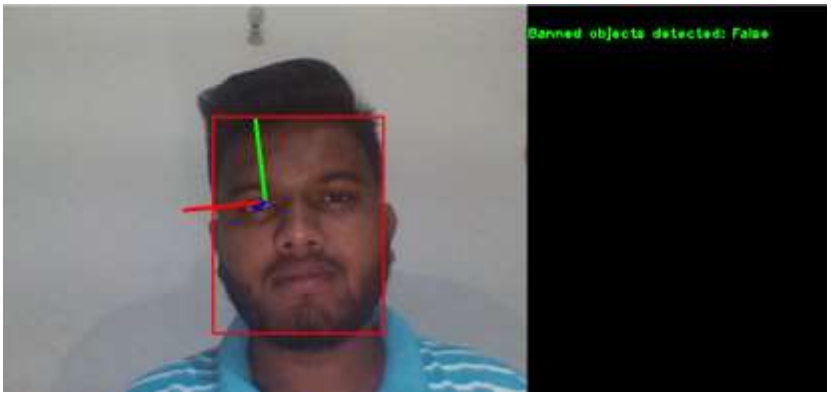

Fig. 6. Output from Object Detection module for frame with no banned objects.

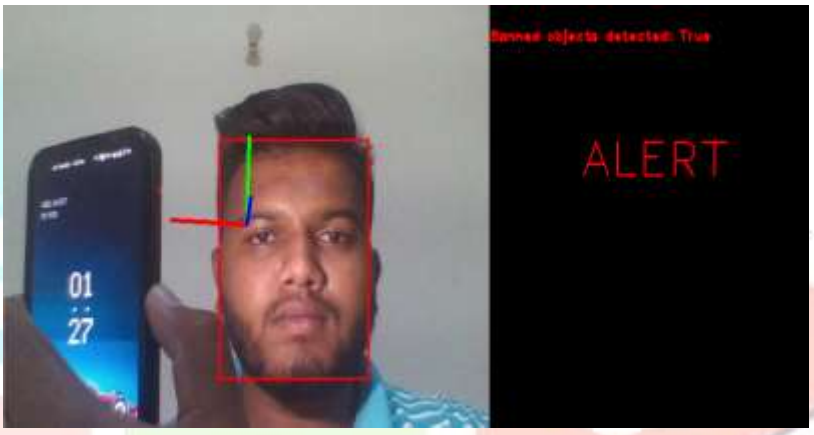

Fig. 7. Output from Object Detection module for frame with banned object.

4.4 Face Detection

We used OpenCV's DNN (Deep Neural Network) module to find the examinee's face in the frame. The face detector is based on the Single Shot Detector (SSD) framework with a ResNet base network.

4.5 Authentication

4.5.1 Face Verification

We used Dlib's [12] face verification model to get the examinee's name. The model uses a pre-trained Resnet50 CNN model to extract a 128D feature vector from all facial images in the database. Then the model uses the same steps on the detected examinee's face to extract a 128D feature vector. After that, Euclidean distance is calculated between the two feature vectors. If the distance is below a certain threshold, then both faces are assumed to be the same. Dlib's default threshold of 0.6 was used for face verification. Outputs from the Face Verification module are illustrated in Fig. 8. and Fig. 9.

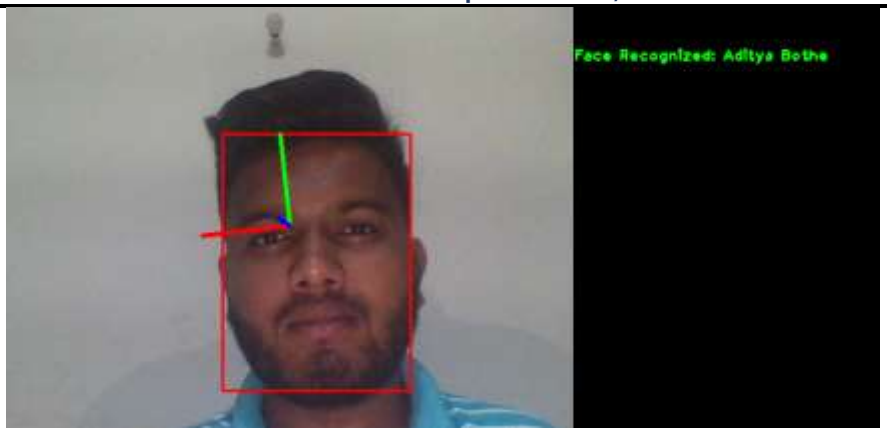

Fig. 8. Output from Face Verification module for frame with valid examinee.

Fig. 9. Output from Face Verification module for frame with invalid examinee.

4.5.1 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 [3] 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. Outputs from the Face Spoofing module are shown in Fig. 10. and Fig. 11.

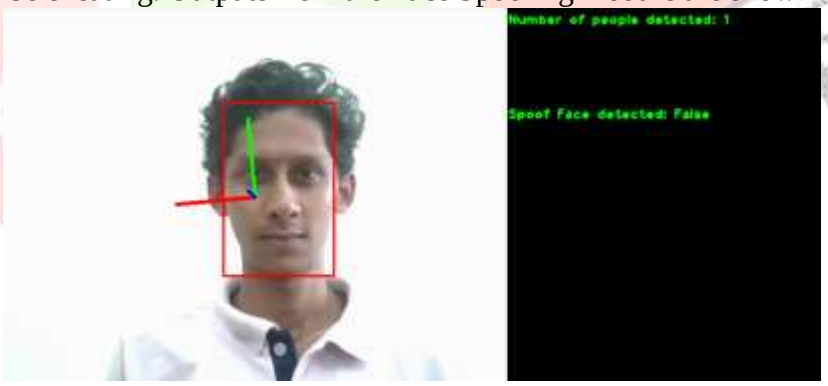

Fig. 10. Output from Face Spoofing module for frame with real face.

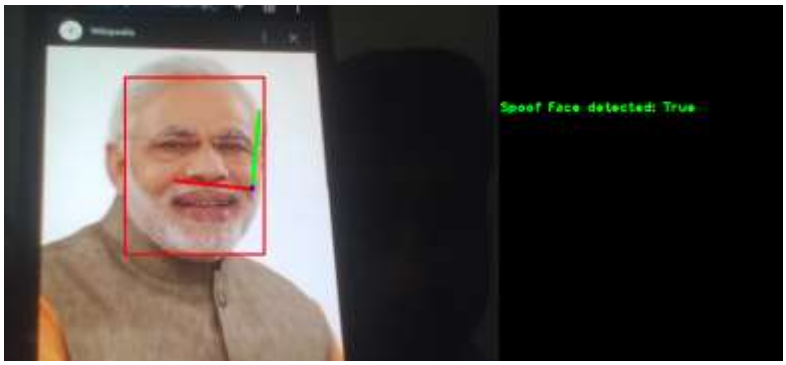

Fig.11 Output with spoofed face

4.6 Abnormal Behavior Detection

4.6.1 Headpose Estimation

We prepared a lightweight head posture estimation demonstrate that can accomplish real-time execution in a framework with moo computational control. Our strategy can foresee precise pitch, yaw, roll points of a individual straightforwardly from the confront edit without the necessity of point of interest or profundity maps. We chosen to prepare a posture estimation organize from scratch instep of utilizing point of interest since we accept that profound systems have expansive focal points compared to landmark-to-pose strategies due to the taking after reasons:

- Deep systems are not subordinate on the head modelchosen, the point of interest location strategy, the subset of focuses utilized for arrangement of the head demonstrate, or the optimization strategy utilized for adjusting 2D to 3D focuses [12].
- They continuously yield a posture forecast which is not thecase for the last mentioned strategy when the point of interest location strategy comes up short particularly for extraordinary postures [12].

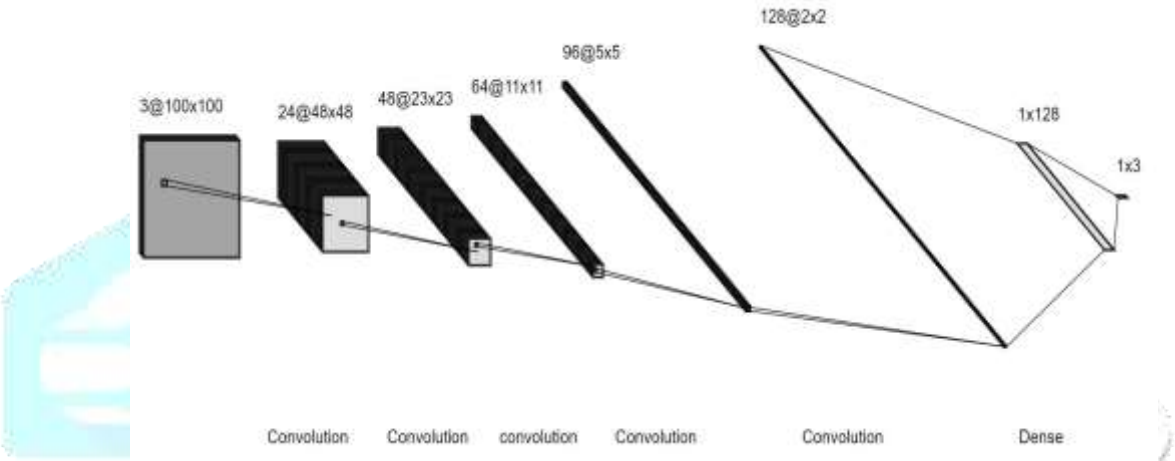

Fig. 12. Convolutional neural network architecture of proposed head pose estimation model.

Our lightweight relapse demonstrate engineering is outlined in Fig. 13. We utilized a 5x5 convolution layer with walk 2 for the to begin with layer taken after by 4 3x3 convolution layers with walk 2 for highlight extraction. The include extraction is taken after by a thick layer with 128 hubs and the yield relapse layer with 3 hubs. For all layers but the final convolution layer and relapse arrange, we utilized the ReLU enactment work. For the final convolution layer some time recently Straighten layer and the taking after thick layers, we utilized the tanh actuation work. At long last, for the yield layer, we utilized the direct actuation work. Fig. 13. outlines the yield from the Head Posture Estimation module.

Fig. 13. Output from Head Pose Estimation module. 3D (red, green and blue) vectors are used to illustrate the predicted head pose angles (pitch, yaw and roll

4.6.2 Eye Tracking

We utilized Dlib's pre-trained organize for recognizing and anticipating 68 facial points of interest on the examinee's confront. Cleared out eye is characterized by the taking after points of interest - 36,37,38,39,40,41. Right eye is characterized by the taking after points of interest - 42,43,44,45,46,47. To begin with, we fragmented the eye locales by utilizing a cover. At that point we connected twofold thresholding on the eye locales to isolated the eyeballs from the rest of the eye locales. Eyeballs gotten to be dark and the rest locales remain white. At that point a vertical separator was made at the centre of each eye. At long last, to decide if the examinee is looking cleared out or right, we characterized an eye-tracking proportion as:

$$AvgETR = \frac{RightEyeETR + LeftEyeETR}{2}$$

where,

$$RightEyeETR = \frac{No.\ of\ white\ pixels\ on\ left\ side}{No.\ of\ white\ pixels\ on\ right\ side}$$

$$LeftEyeETR = \frac{No.\ of\ white\ pixels\ on\ left\ side}{No.\ of\ white\ pixels\ on\ right\ side}$$

After broad trial and testing, we settled the taking after limits for the AvgETR: $\leqslant$0.35 (looking exterior the screen), 0.36 to 3.9 for the center (looking at the screen), $\geqslant$4 for cleared out (looking exterior screen). If the examinee is looking exterior the screen for more than 10 continuous outlines, at that point the examinee is said to be cheating. Yield from the Eye-tracking module is appeared in Fig. 14.

Fig. 14. Output from Eye tracking module

4.6.3 Mouth Movement Analysis

Lip locale is characterized by the taking after points of interest - 60, 61, 62, 63, 64, 65, 66, 67. To decide if the mouth is open or closed, we characterized a lip angle proportion as:

$$L.A.R = \frac{|P(62)-P(66)|}{|P(60)-P(64)|}$$

After broad trial and testing, we settled the limit to be 0.1. If L.A.R>0.1, this implies the mouth is open, else it is closed.

Furthermore, to check if the individual is talking or not, we characterized a buffer. If the examinee keeps his mouth open for more than 10 successive outlines, at that point the action is classified as talking and subsequently cheating. Yields from the Mouth Development Investigation module are appeared in Fig. 15. and Fig.16.

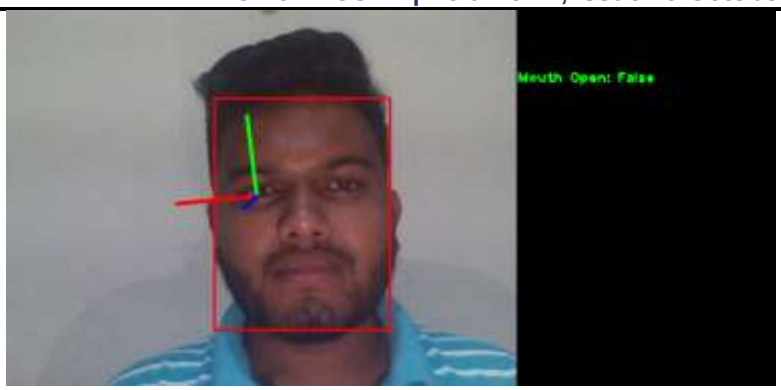

Fig. 15. Output from Mouth Movement Analysis module for frame containing examinee with closed mouth.

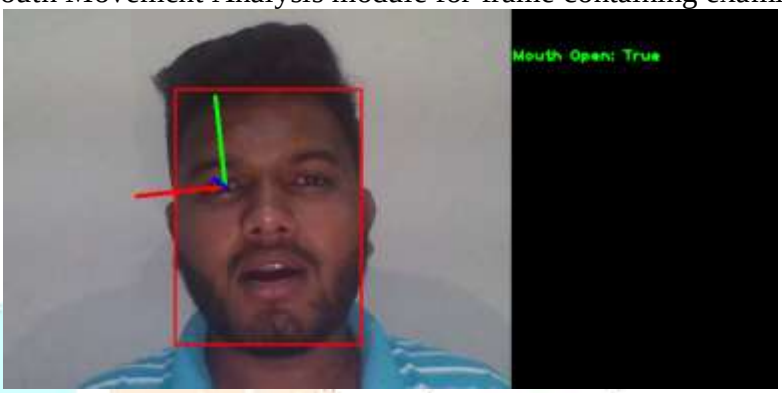

Fig. 16. Output from Mouth Movement Analysis module for frame containing examinee with open mouth.

V. EXPERIMENTAL RESULTS

5.1 Head Pose Training

Our demonstrate was prepared on pictures from the Pandora dataset. The show takes a confront trim rescaled to 100 x 100 as input. The confront discovery bounding box yield is extended by 100% and the coming about bounding box is utilized to edit out the head locale and pass that as input to the organize. We utilized the Adam optimizer with a learning rate of 0.001 to prepare the demonstrate. The demonstrate was either prepared for 100 ages or was utilized with early ceasing which observed the approval misfortune with the persistence of 20 ages. The comparing preparing misfortune and approval misfortune accomplished for the best show is appeared in Fig. 17. The demonstrate was prepared utilizing Cruel Squared Mistake (MSE) as a misfortune work and assessed on the test dataset utilizing Cruel Normal Blunder (MAE). To avoid the demonstrate from overfitting we utilized dropout regularization after each convolution and thick layer.

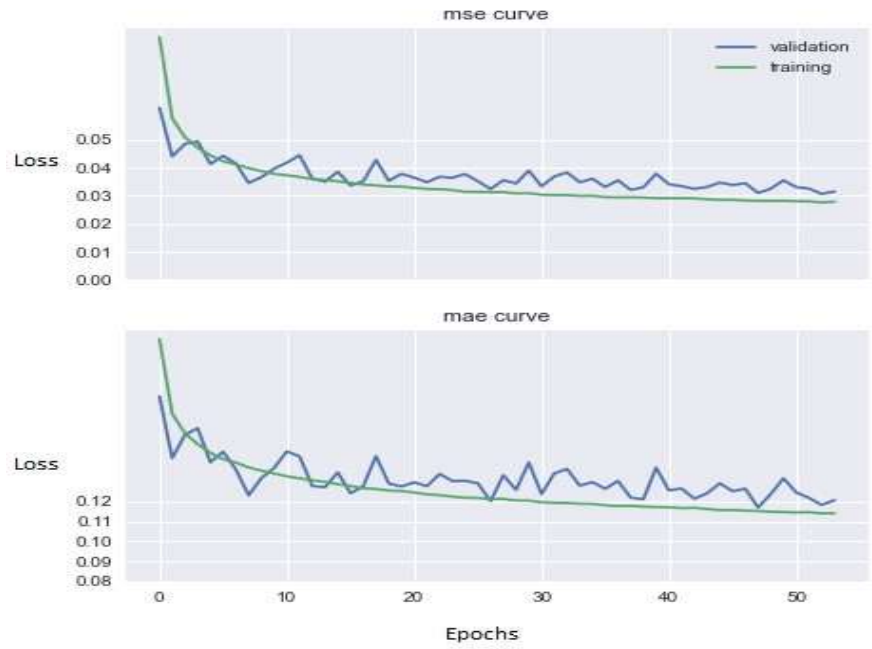

Fig. 17. Training and validation loss curves for Head Pose Estimation model.

5.2 Head pose Evaluation

We assessed our lightweight head posture estimation demonstrate on the BIWI benchmark dataset. The dataset has around 15k RGB pictures with ground truth head posture point values. The proprietors of the Pandora dataset have moreover given the trimmed confront locale for the BIWI dataset. So, we utilized this trim straightforwardly without physically trimming the confront locale from the unique benchmark dataset. Our lightweight head postures estimation demonstrate was able to accomplish a superior precision compared to celebrated landmark-based approaches and 3D thick modelbased strategies. Table 3 appears the comparison of comes about gotten by our demonstrate with Dlib [12] and 3DFFA [6] models. We took execution assessment comes about of Dlib and 3DFFA from Hopenet's (state-of-the-art show) paper. Hopenet is a profound learning-based classification strategy for fine-grained head posture estimation. Indeed in spite of the fact that the exactness gotten by Hopenet is much higher than our lightweight show, we haven't utilized it in our online proctoring examination framework. The Hopenet employments Resnet as the spine include extractor since of which, the computational complexity of the demonstrate is much higher and will not give real-time execution for low-cost frameworks. Consequently its execution is not compared with our lightweight demonstrate in Table 3. MAE is utilized to assess the execution of all models.

Fig. 18. outlines the anticipated head posture vectors of our demonstrate on the BIWI benchmark dataset. This appears that our show was able to perform nicely indeed for expansive head posture angles.

Table 3. Comparison of proposed head pose estimation model accuracy (MAE) with other available models

Model	Pitch	Yaw	Roll	MAE
Our model	11.000	13.927	7.471	10.799
Dlib	13.802	16.756	6.190	12.249
3DFFA	12.252	36.175	8.776	19.068

Fig. 18. Results by proposed head pose estimation model on BIWI benchmark dataset.

VI. Conclusion and Future Work

In this work, we proposed a Cleverly Framework that employments a webcam to screen examinees amid the online examination. The arrangement offers a comprehensive checking and investigation suite to anticipate examinees from cheating in an online exam. Functionalities incorporate client confirmation and unusual behavior monitoring.

Currently, our pipeline runs at less than 30 Frames per second (fps) since of the complexity of the models that we have utilized. In the future, we are arranging to move forward the fps by utilizing lightweight models and appropriate optimization methods like quantization. Besides, the execution of the confront spoofing show is not palatable. In the future, we are arranging to supplant it with a more exact show. As of now, we are utilizing a picture handling-based eye-tracking strategy. Afterward, we'll supplant it with a profound learning-based look estimation show that precisely gauges where the examinee is looking.

VII. 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. 1867– 1874.

[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, "Face-from-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.

[11] Gary Bradski and Adrian Kaehler, "Opencv," *Dr. Dobb's journal of software tools*, vol. 3, 2000.

[12] S Sharma, Karthikeyan Shanmugasundaram, and Sathees Kumar Ramasamy, "Farec—cnn based efficient face recognition technique using dlib," in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*. IEEE, 2016, pp. 192–195.

[13] Nataniel Ruiz, Eunji Chong, and James M Rehg, "Finegrained head pose estimation without keypoints," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 2074– 2083.