An automated proctoring assistant in online exams using computer vision

Nguyen Khanh Luan, Pham Thi Thu Ha

Instructor

Dr. Phan Duy Hung

TABLE OF CONTENTS

I. INTRODUCTION

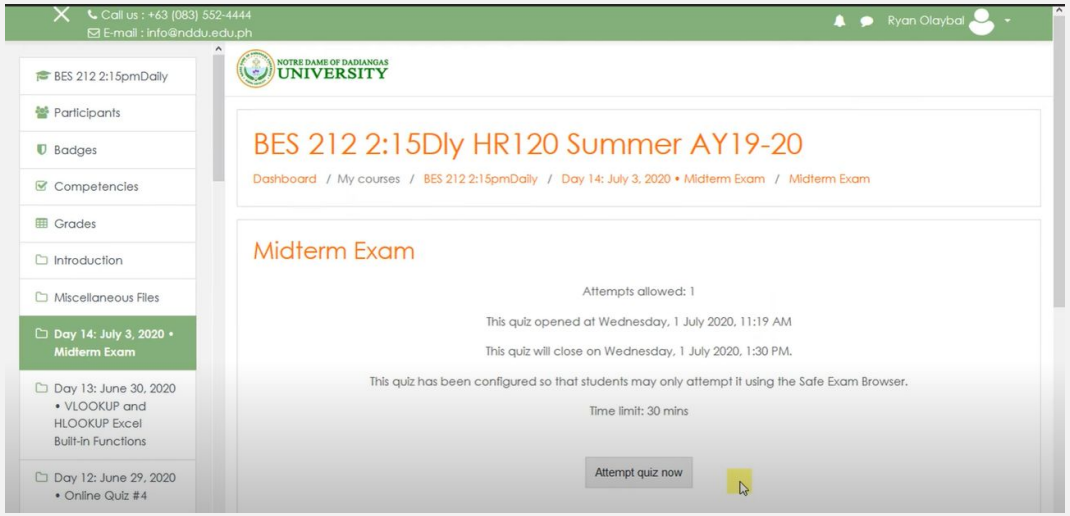
II. METHODOLOGY

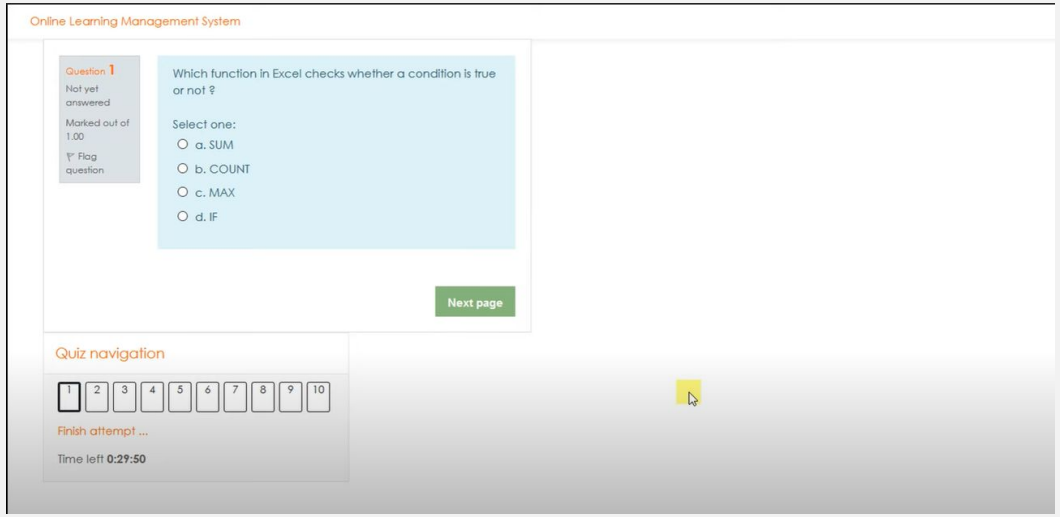
III. EXPERIMENTAL RESULTS AND DISCUSSIONS

IV. CONCLUSION AND FUTURE WORKS

I. INTRODUCTION

- Amid the COVID-19 pandemic, most schools and colleges have turned to online learning and testing in response to the long-term quarantine.
- Cheating or attempting to cheat in education has had the opportunity to become numerous and complex since the outbreak of the COVID-19 pandemic.
- With the online exam, many methods have been implemented, such as monitoring candidates via Zoom online video and audio, studying online and taking the test offline under the supervision of a proctor, or taking the online exam on a secure browser.



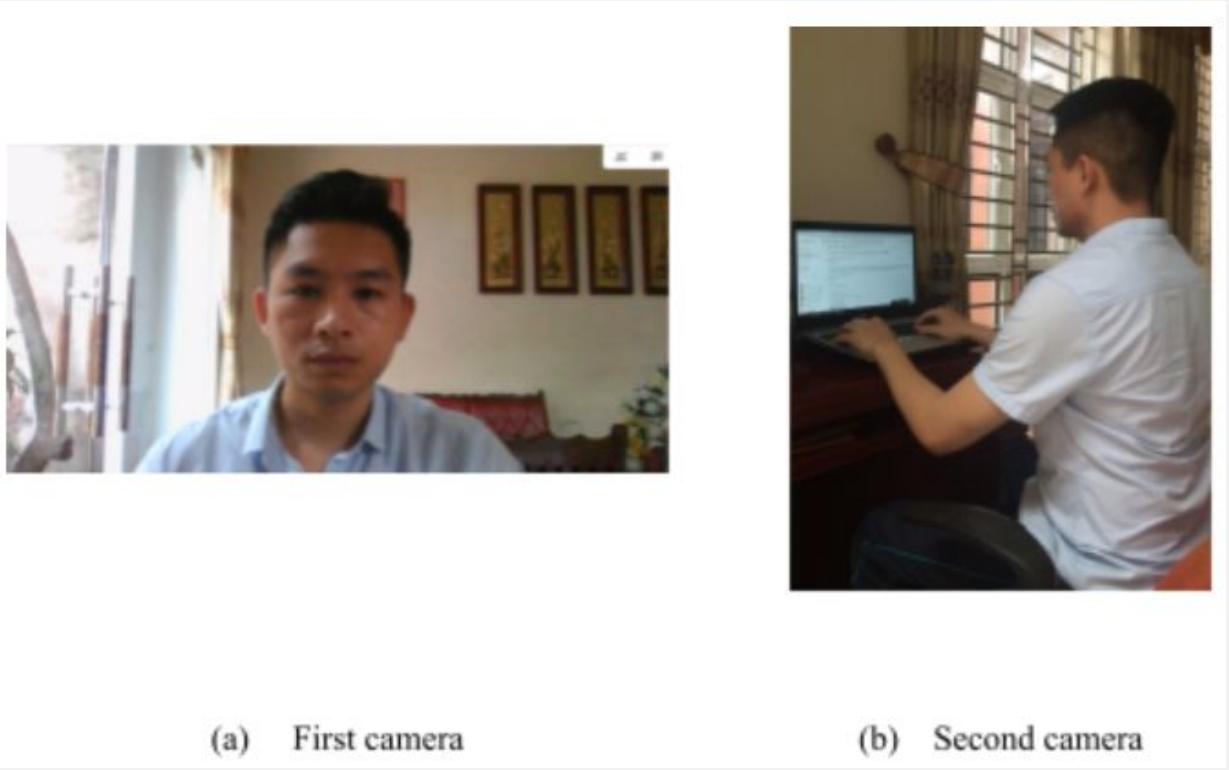


Take the remote test on Safe Exam Browser - a secure browser

- Most learners today are capable of equipping personal computers and mobile phones.

- A reliable method requires students to set up two cameras during the exam:
- The first camera allows observing the front of a student
- The second camera can show the surrounding environment

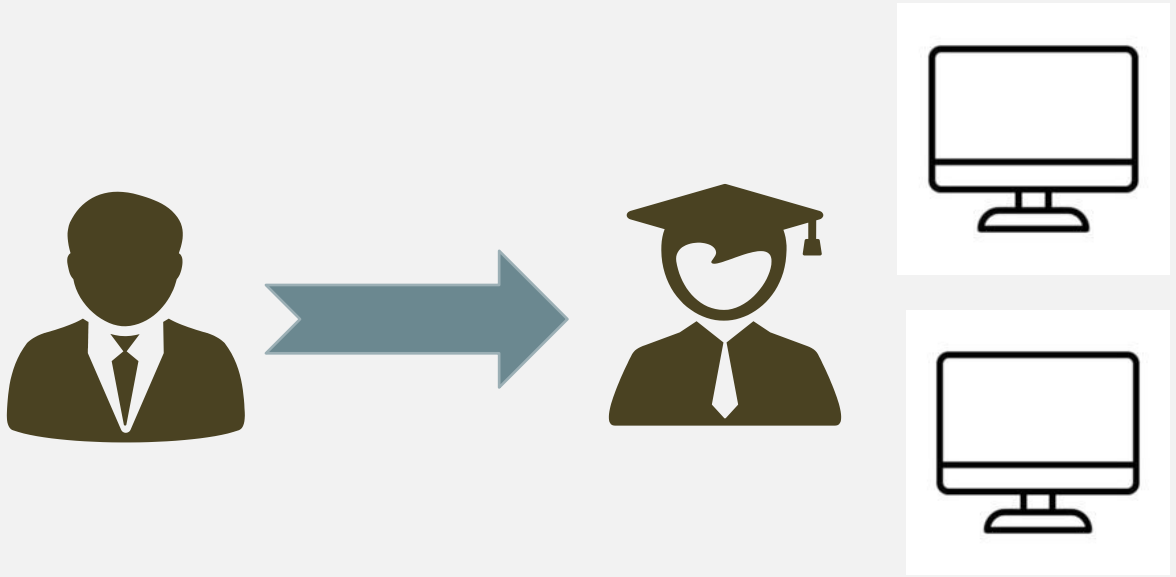
However, this method still needs the supervision of the supervisors via the cameras.



Arrangement of two cameras to capture students' images

Problem

- Fall 2021 semester - Hoa Lac campus: 116 subjects
- JDP113 subject:1020 students were taking the test. If each online exam room has > 20 students -> 51 online exam rooms
- If each exam room has 1 proctor, each proctor has to simultaneously observe about 40 screens through the cameras



-> organizing online exams require a massive amount of effort, time, and money to ensure the fairness of the assessment

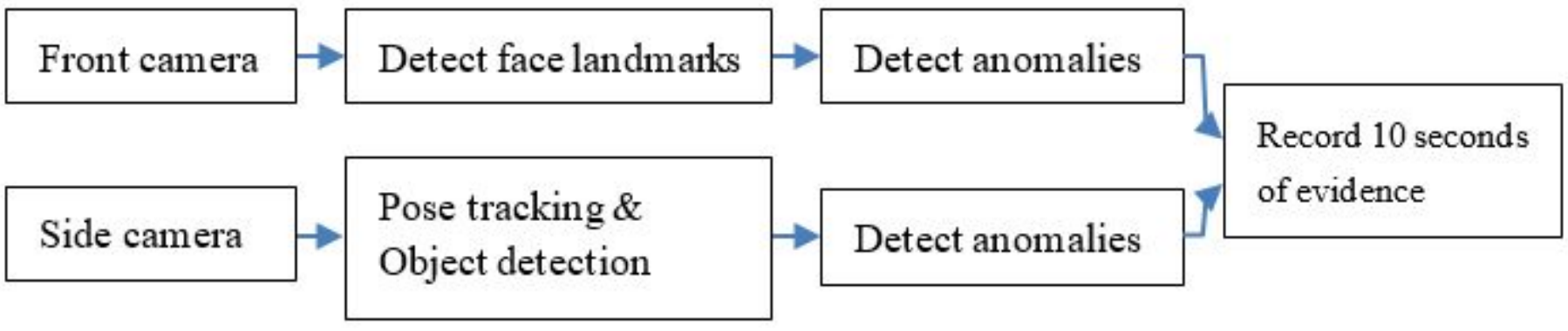
This work implements a computer vision system to observe and record unusual behaviors to assist online exam proctoring and reduce the proctor's effort

Related Works

Søgaard	Safe Exam Browser ▪ host a user system and shut down unrelated apps or websites ▪ disable screen recordings and projection ▪ participants can still access outside materials
Bedford et al	▪ detects devices under usage, searches for similar test contents, and flags certain activities ▪ scan the internet, block the sites ▪ participants can still attempt to cheat by using offline resources
Sahil et al	▪ system asks students to provide their ID, photo, and personal information during the registration process ▪ verify legitimate candidates using face recognition ▪ needs multiple checks
Turani et al	▪ a proctoring system using a 360-degree security camera ▪ excellent audio and video qualities ▪ Expensive, hard to wear

Contribution

Proposed System



An automated AI-based proctoring system:

- The input images are fed from two cameras
- Front camera: face landmark model to track suspicious head, eyes, lips movement
- Side camera: a pose recognition model based on a long short-term memory, object detection

II. METHODOLOGY

Problem assumptions

- Students need to sit in front of the computer during the test-taking; other moving or cheating behavior are not allowed.
- Cameras are set up to clearly see the faces of test-takers, their bodies, and the surrounding environments.
- Students are not allowed to bring electronic devices that can receive or transmit information, such as mobile phones, USBs, and memory cards.
- Students can only use pens, white paper, or material approved by the official proctor.
- Other than the test taker, no one is allowed to enter the room, including friends or family members.

Supervising front camera

Eyes tracking

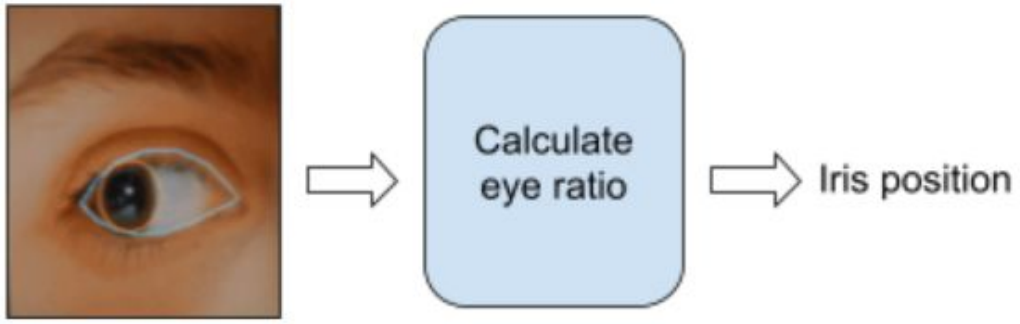


Figure 6. Iris tracking approach

- Using landmarks coordinates output from Mediapipe Iris
- A suspicious glance is counted as the ratio > 1/3
- Eye width and distance from the iris to the eye edge are calculated

Head pose estimation

- A combination of Mediapipe 3D face mesh and OpenCV Perspective-n-Point (PnP)
- Given a set of 3D points in the world and their corresponding 2D projections in the image, PnP can estimate the pose of a calibrated camera

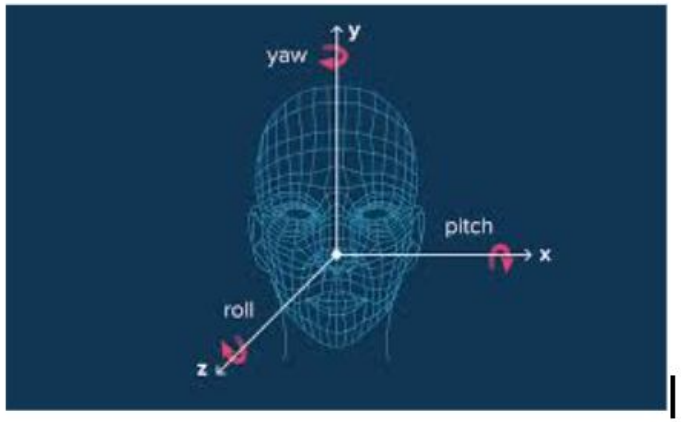


Figure 8. *Angles are represented by different axis rotation*

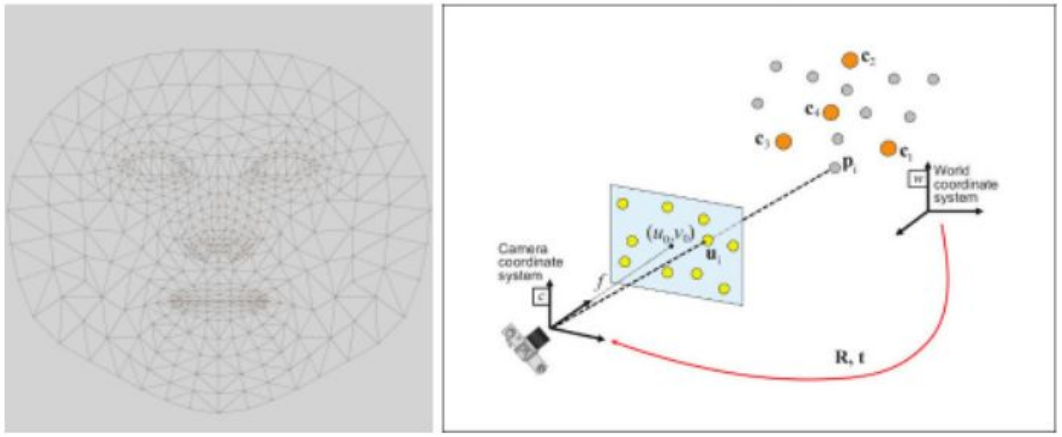
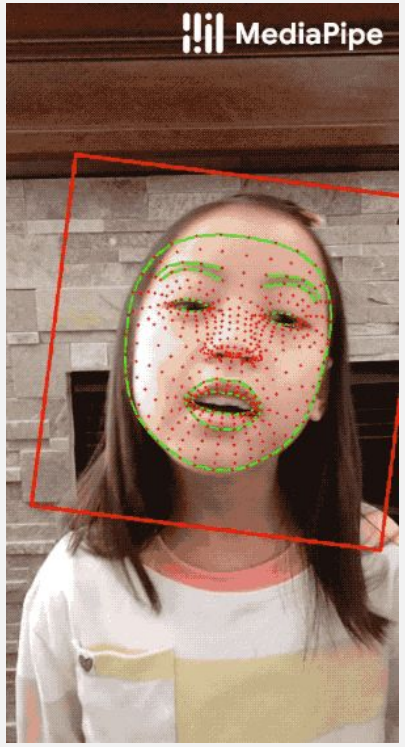
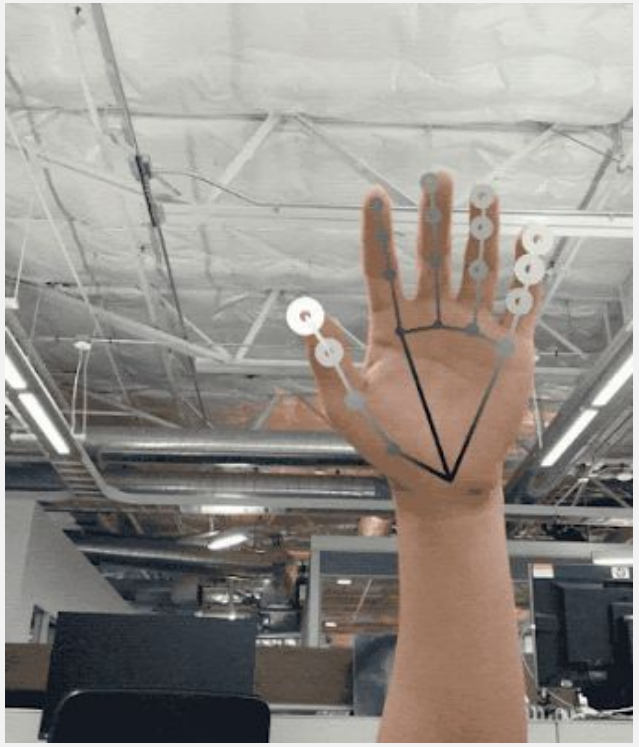


Figure 7. *Face mesh Mediapipe and PnP OpenCV*

Mouth and hand tracking

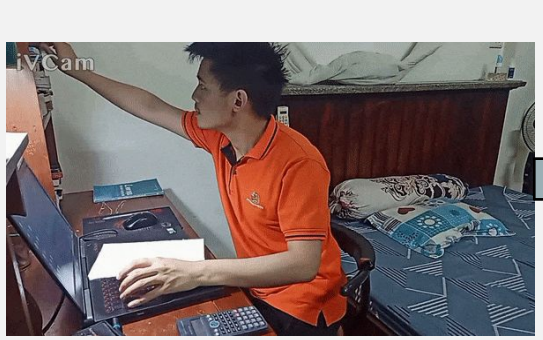


- A student is considered talking when the distance between his or her lips is greater than a predefined distance and over a period of time
- Efforts to occlude the mouth using hands are also dealt with by making sure mouth coordinates do not fall into hand area coordinates



Supervising side camera

Pose tracking and detecting suspicious actions



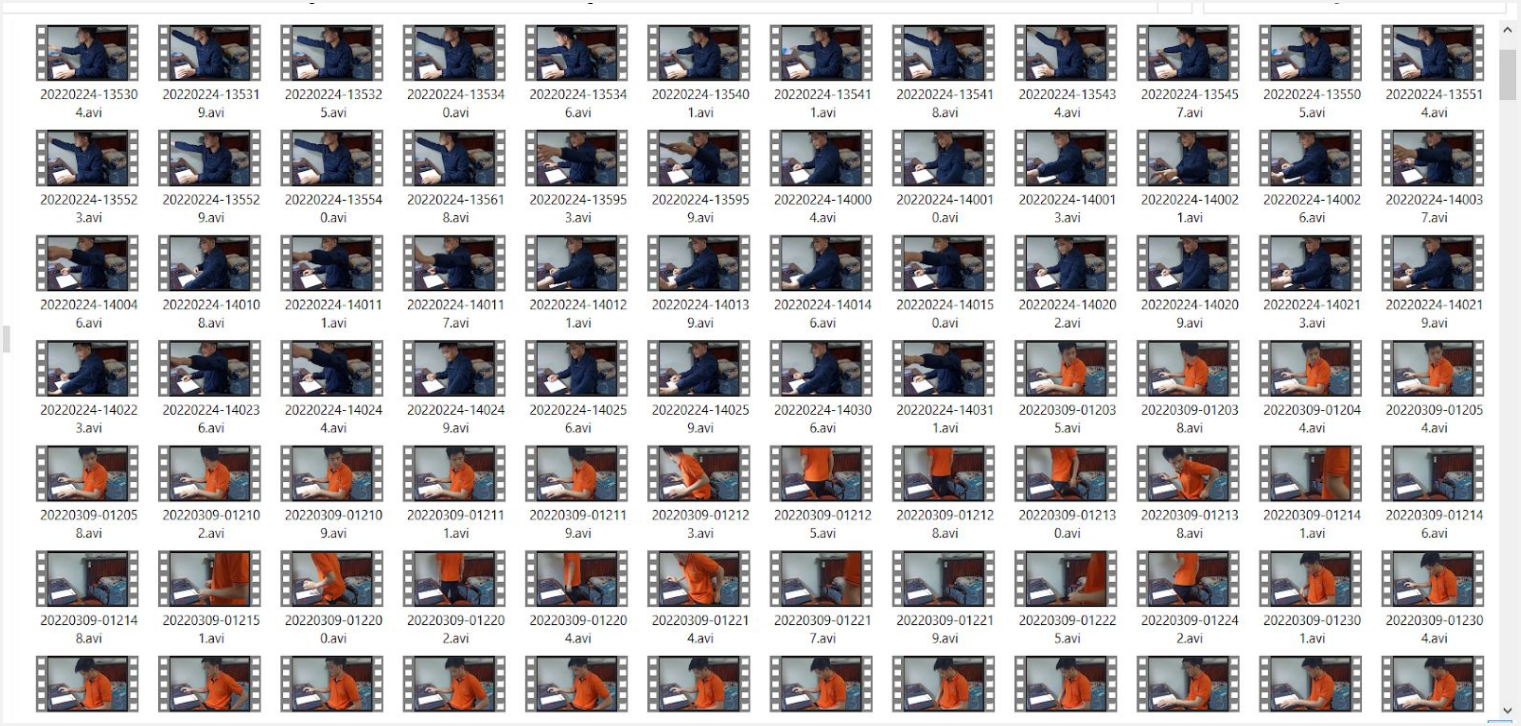
Pose Detector

LSTM Networks

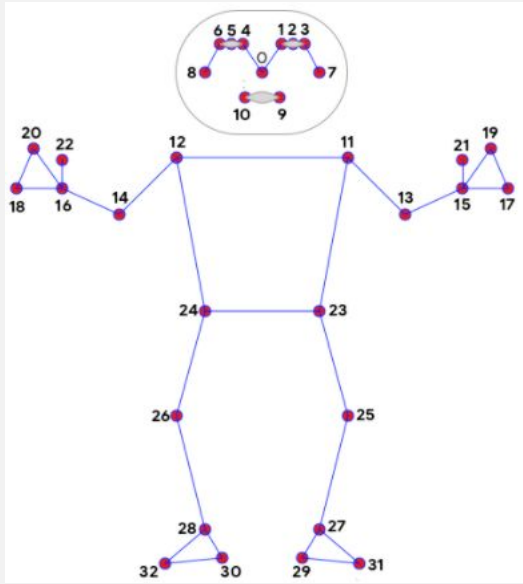
SUSPICIOUS

Data collection

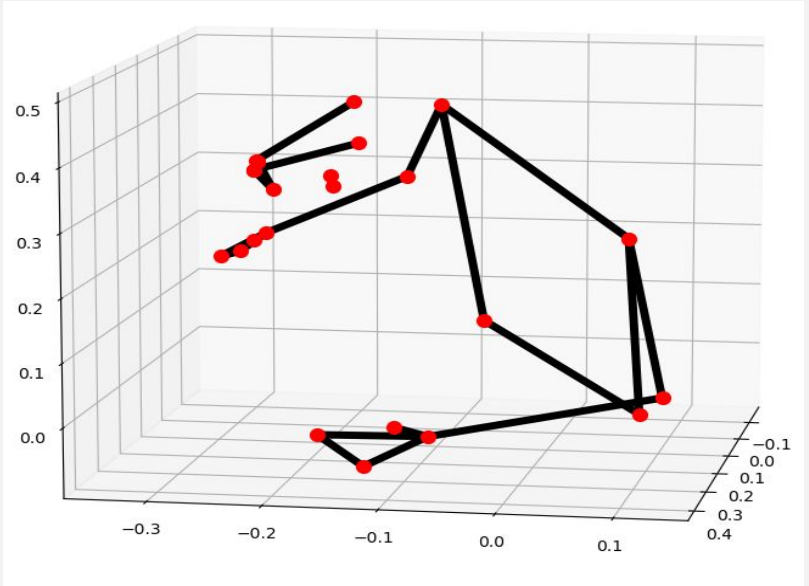
- *Recording and labeling video dataset*
- A batch of 720 videos of 1 second long is labeled as cheating and 720 videos is labeled as non-cheating
- Mediapipe Pose



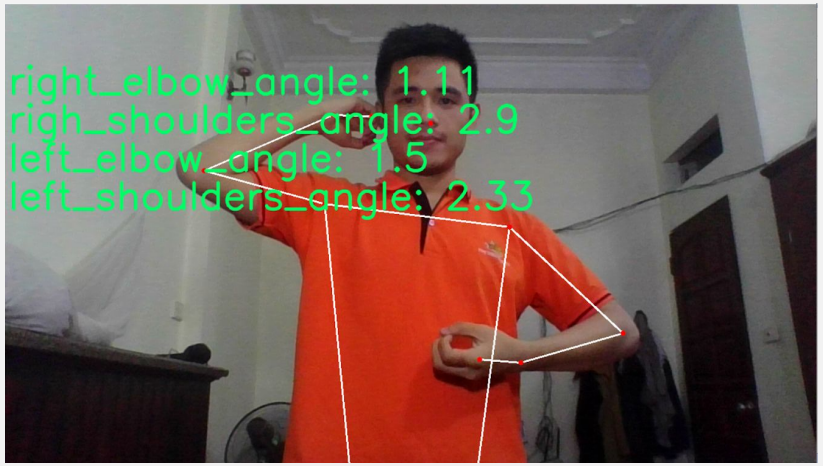
Pose landmarks



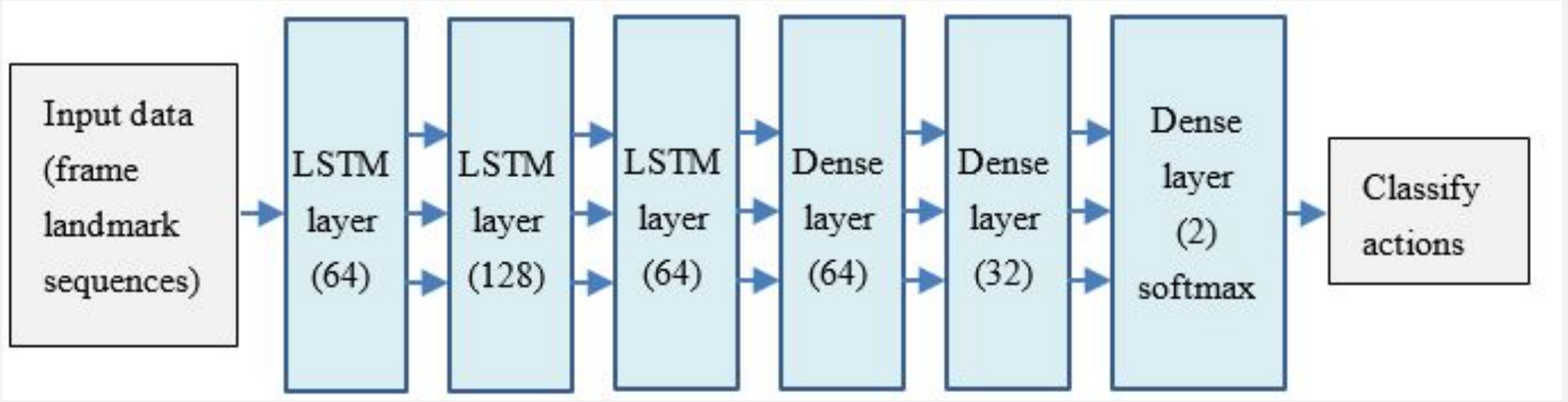
0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index



3D landmarks coordinates and angles extractions of student's elbow and shoulder

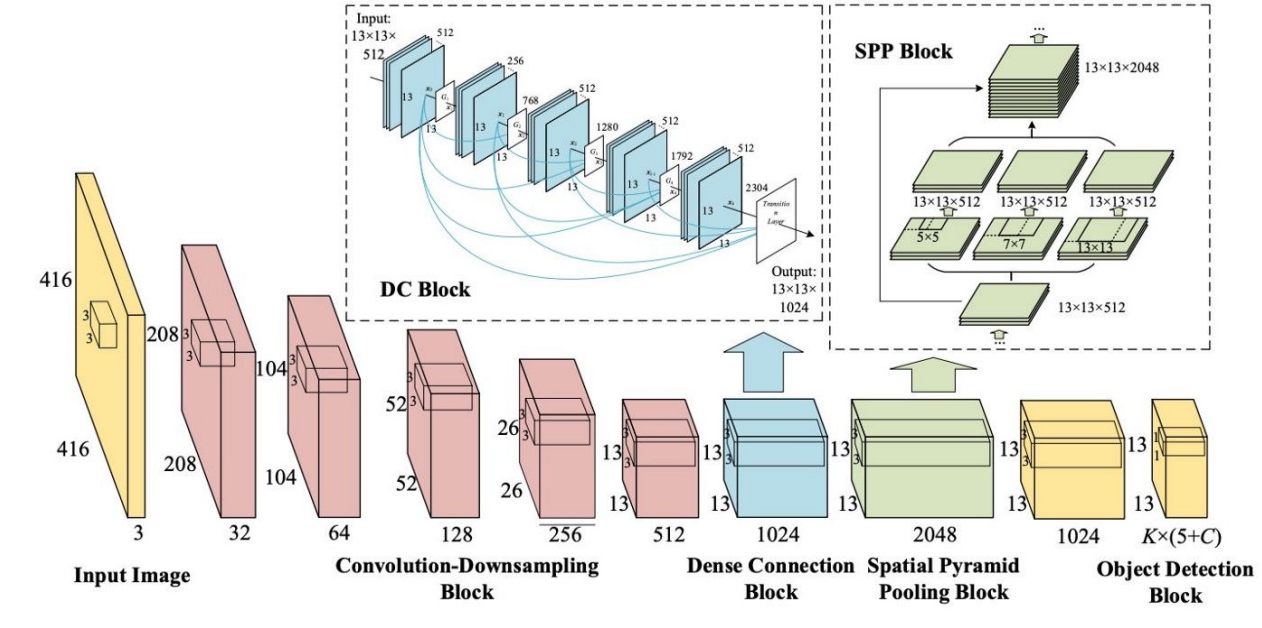


Pose recognition model



LSTM network architecture for classifying action

Object detection and people counting



YOLOv4 network structure

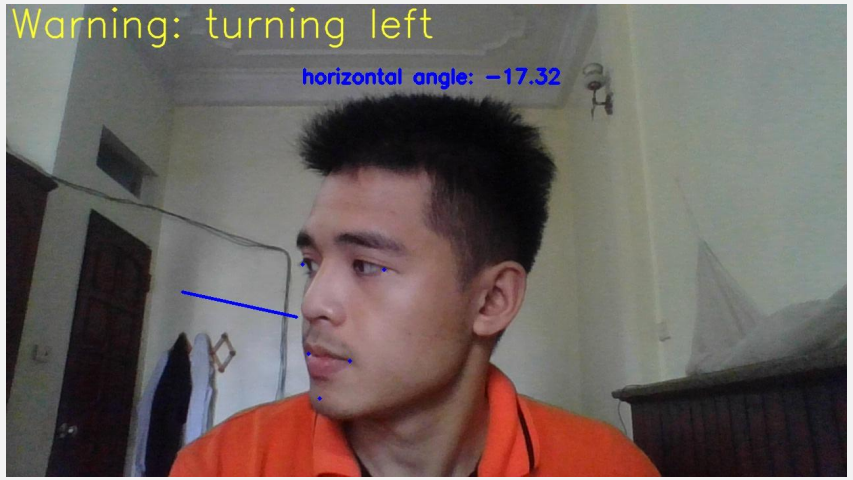
- The custom YOLOv4 model applies a pre-trained model with Alex's Darknet to recognize 8 types of objects for this specific case - **8 classes** including: person, tv monitor, laptop, mouse, remote, keyboard, cell phone, book

- The model will detect, draw predicted bounding boxes and count the number of objects of each class with confidence

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

Front camera results

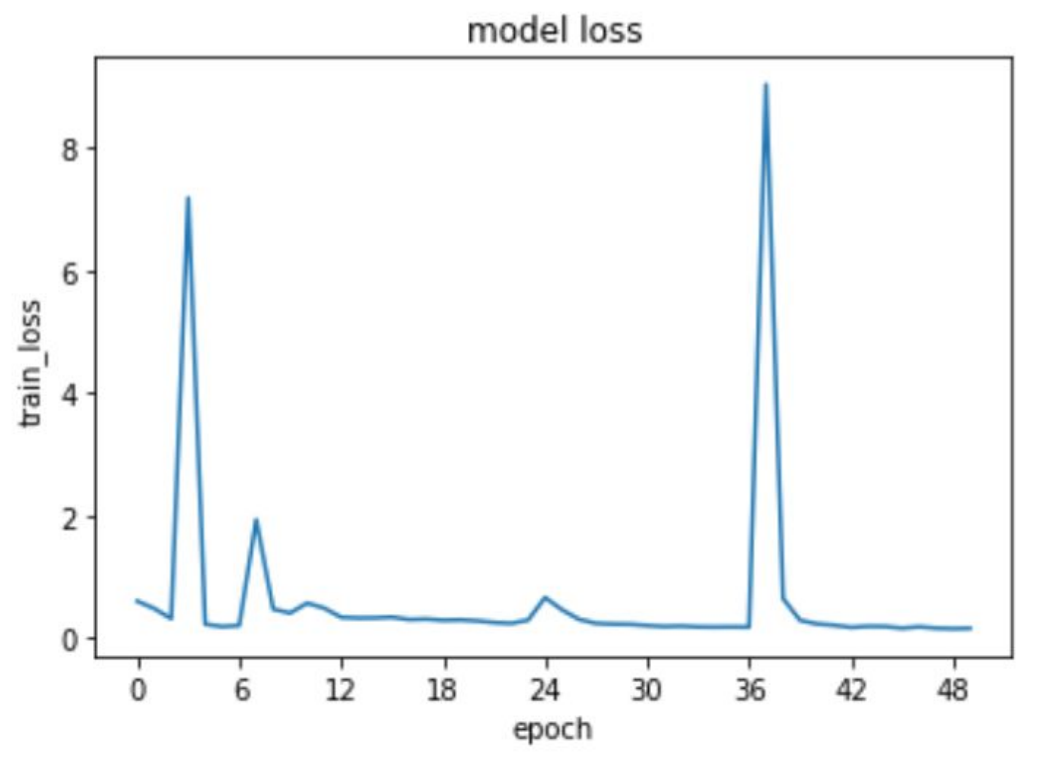
Eyes tracking result

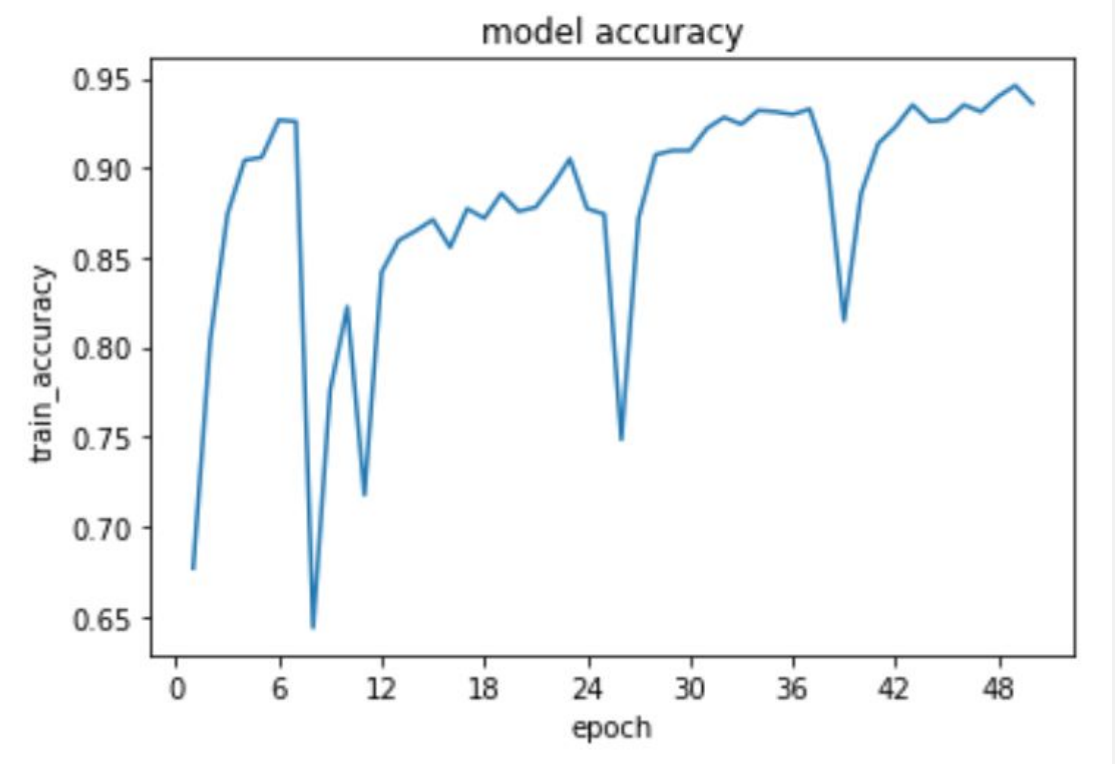


Head pose estimation

Mouth tracking and face occlusion detected

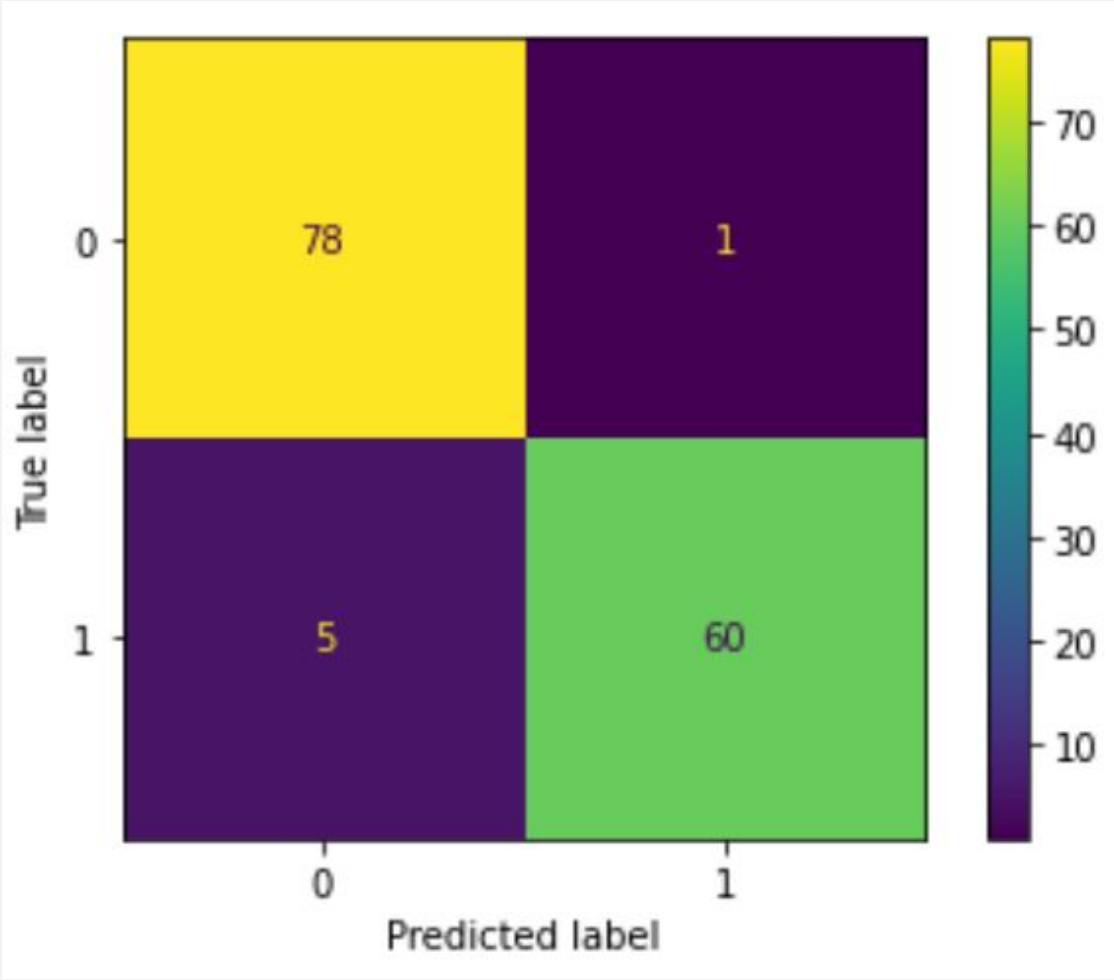
Side camera results

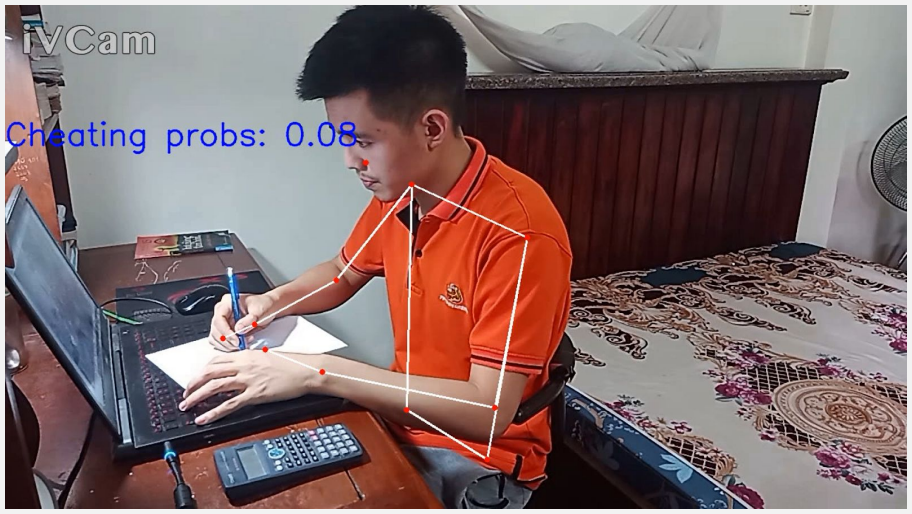




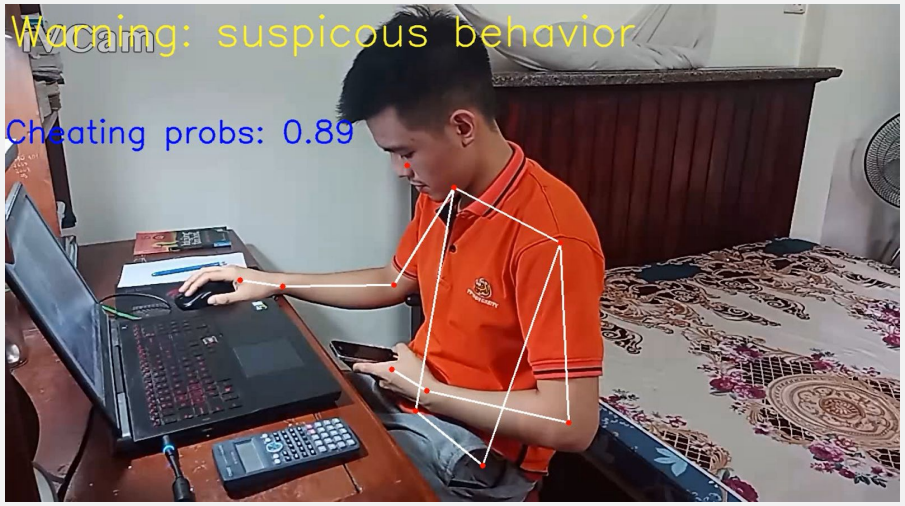
LSTM Model loss and accuracy on training dataset

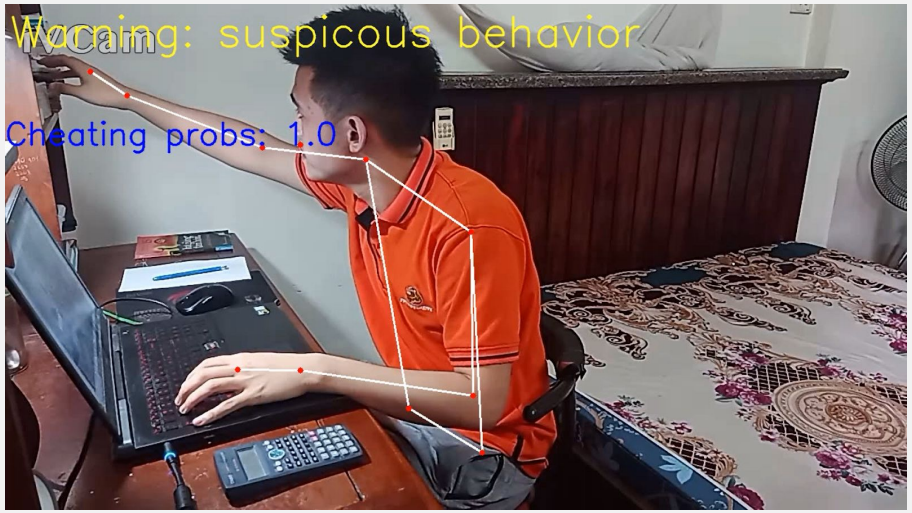
Confusion matrix to evaluate model accuracy





No suspicious action was detected

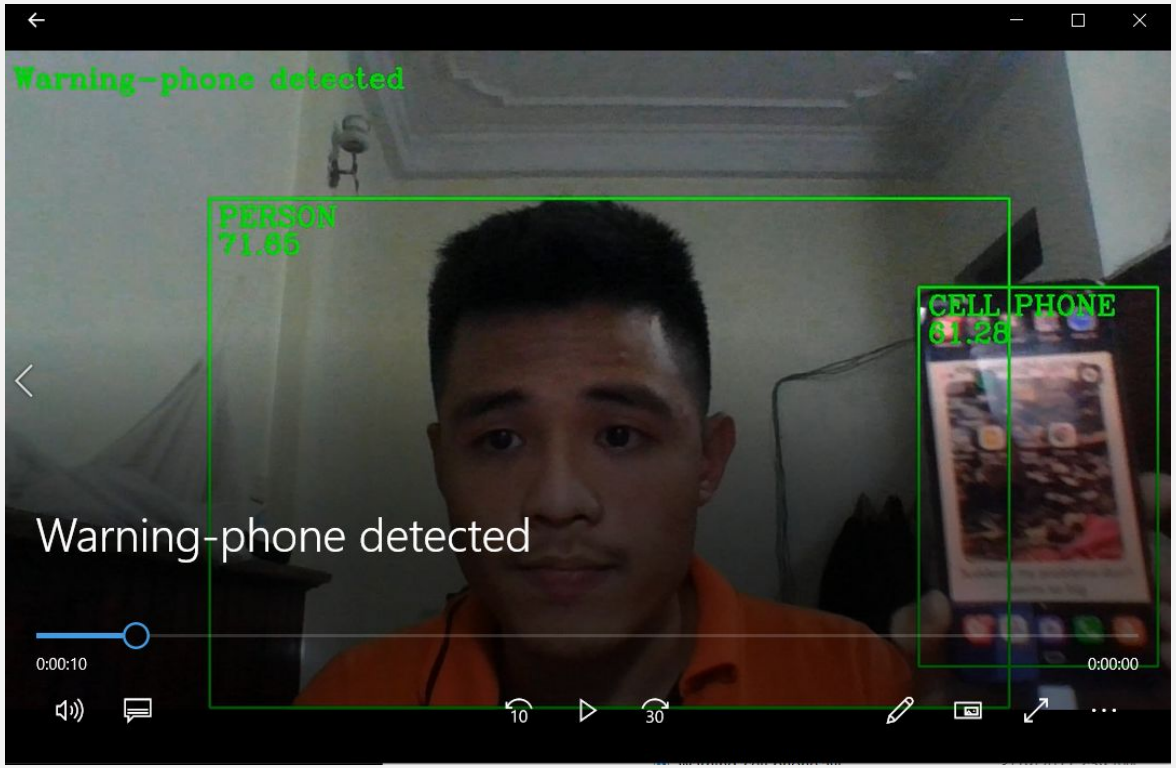




Suspicious action detected

Results of object detection and person counting

As a student is flagged with any suspicious action, this system automatically logs them back for a 10-second long duration with a corresponding name



Video logged by the system

IV. CONCLUSION AND FUTURE WORKS

Conclusion

- This thesis proposes a proctor assistant system that can detect students' anomaly actions through cameras and record them using computer vision
- This system shows promising results and can be used to aid teachers and proctors when monitoring the examination
- The system saves a lot of time and effort for the organization

Future Works

- Many improvements can be made based on the proposed model including:
  - automatic identification of student identities based on the university's face database
  - customizing object recognition to the specific requirements of the various exams
- The results continue to be improved as more new real-life situations are added

QUESTIONS

THANK YOU FOR LISTENING !