

GIẢI PHÁP HỆ THỐNG PHÁT HIỆN GIAN LẬN TRONG PHÒNG THI

ANTI-EXAM CHEATING DETECTION SYSTEM

Đoàn Thị Hương Giang^{1*}, Hồ Đoàn Bảo Châu²

DOI: <http://doi.org/10.57001/huinh5804.2025.056>

TÓM TẮT

Khoa học công nghệ ngày càng phát triển, việc áp dụng công nghệ Tự động hóa, Điện tử, IoT và Trí tuệ nhân tạo vào các ứng dụng thực tế được thực thi trên nhiều lĩnh vực khác nhau của đời sống hiện đại. Trong đó, áp dụng công nghệ trong quản lý hoạt động trong giáo dục ngày càng được chú trọng. Đặc biệt là quản lý gian lận tại các kỳ thi luôn được xem là một vấn đề cấp bách, tốn nhiều nhân lực cũng như không thể tránh khỏi việc bỏ sót khi số lượng người thực hiện giám sát trong mỗi phòng thi có hạn. Nghiên cứu này đề xuất giải pháp sử dụng kiến thức liên ngành nhằm xây dựng mô hình giám sát hành vi gian lận tại phòng thi trắc nghiệm trên máy tính. Hệ thống đề xuất được thực hiện với hai tầng giám sát gồm: (1) Hệ thống phát hiện gian lận tại cửa phòng thi sử dụng hai lớp kiểm tra bằng thẻ từ và nhận dạng khuôn mặt; (2) Hệ thống phát hiện hành vi gian lận trong phòng thi. Trong đó, việc phát hiện gian lận trong phòng thi chúng tôi định nghĩa, xây dựng, thu thập và gán nhãn một bộ cơ sở dữ liệu hành vi gian lận bằng hình ảnh. Sau đó, mô hình phát hiện hành vi gian lận với một hàm mục tiêu mới cho kiến trúc YOLO V5 được triển khai. Mô hình này tăng hiệu quả của quá trình phát hiện gian lận trong phòng thi trắc nghiệm bằng hình ảnh. Giải pháp đề xuất cho kết quả đạt độ chính xác cao hơn mô hình YOLO V5 gốc từ 7,2% đến 9,88%.

Từ khóa: Mạng nơ ron tích chập, học sâu, gian lận thi cử, nhận dạng hành động bất thường, phát hiện hành động.

ABSTRACT

Science and technology are increasingly developing, the applications of Automation, Electronics, IoT, and Artificial Intelligence technologies into practical applications have been applied in many different areas of our modern life. Including management of activities in education. In particular, managing cheating in exams is always considered an urgent problem, consuming a lot of human resources as well as inevitably missing when the number of people performing supervision in each room is limited. This research proposes a solution that utilizes these interdisciplinary knowledges to build a cheating behavior detection system in computer-based examination rooms. The proposed system is implemented with two monitoring layers including: (1) Cheating detection system at the examination room door using two cascade layers of checking by RFID card and an online facial recognition; (2) Cheating detection system in the examination room. In order to detect cheating in the exam room, we first define, build, collect and anotate a cheating image database. Then, we use YOLO V5 model with a new objective function to increase the efficiency of the cheating action detection. Our proposed method gives a higher results than the original YOLO V5 model on our EPUExCheating from 7.2% to 9.88%.

Keywords: Convolution neuron network, deep learning, exam cheating, abnormal action recognition, abnormal action detection.

¹Khoa Điều khiển và Tự động hóa, Đại học Điện lực

²Trường Công nghệ thông tin và Truyền thông, Đại học Bách khoa Hà Nội

*Email: giangdth@epu.edu.vn

Ngày nhận bài: 12/12/2024

Ngày nhận bài sửa sau phản biện: 25/02/2025

Ngày chấp nhận đăng: 28/3/2025

1. GIỚI THIỆU

Ngày nay, nhu cầu học tập của người dân ngày càng được chú trọng. Cùng với đó là có nhiều kỳ thi được tổ chức để kiểm tra và đánh giá năng lực của người học. Các hình thức thi ngày càng được đa dạng hóa dưới nhiều hình thức khác nhau như thi tự luận, thi trắc nghiệm trên giấy, thi trắc nghiệm trên máy,... Trong số các hình thức thi đó thì kỳ thi trắc nghiệm trên máy tính ngày càng chứng tỏ được ưu điểm cho cả người thi và người làm công tác quản lý như nâng cao năng lực sử dụng công nghệ thông tin, chấm điểm một cách tự động, thí sinh có thể biết điểm ngay sau khi kết thúc bài thi, số lượng câu nhiều sẽ tăng cường tính bao phủ chương trình đào tạo, dễ dàng trộn và tạo nhiều mã để khác nhau,... Do đó, hình thức thi trắc nghiệm trên máy được sử dụng để đánh giá không chỉ trong các kỳ thi đánh giá năng lực đầu vào của các trường Đại học, Cao đẳng, kỳ thi đánh giá năng lực ngoại ngữ mà cũng được sử

dụng rất nhiều trong các kỳ thi kết thúc học phần tại các trường Đại học, Cao đẳng. Tuy nhiên, hình thức thi trắc nghiệm trên máy hiện nay có một số đặc điểm hạn chế chính như sau: (1) Giám thị cần kiểm tra thẻ sinh viên bằng mắt thường để so sánh với người thật. Các ảnh thẻ thường đã được chụp một cách cố định và cách thời điểm thi khá lâu nên dẫn tới việc kiểm tra khó khăn thậm chí khiến giám thị có thể đối chiếu sai. Để khắc phục hạn chế của các cách quản lý truyền thống cũng như đưa ra giải pháp toàn diện cho quá trình quản lý phòng thi. Trong nghiên cứu này chúng tôi sẽ xây dựng hệ thống giám sát, đối chiếu một cách tự động ngay từ cửa thông qua hai lớp bảo mật là thông qua thẻ từ và nhận dạng khuôn mặt trực tuyến từ hình ảnh thu nhận của camera; (2) Ngoài ra, trong phòng thi trắc nghiệm thường có tường chắn giữa các máy tính với nhau, tránh các thí sinh trao đổi và nhìn bài nhau. Do vậy, kiến trúc này gây khó khăn trong việc bao quát, quan sát và phát hiện gian lận của giám thị khi giám sát trực tiếp. Hoặc nếu có trang bị hệ thống camera quanh phòng thi thì giám thị cần đảo mắt liên tục giữa các không gian camera, trong khi mỗi vùng quan sát của thí sinh trên ảnh là rất nhỏ. Gây khó khăn hoặc bỏ sót cho việc phát hiện hành vi gian lận của thí sinh trong phòng thi. Nghiên cứu này đưa ra giải pháp phát hiện hành vi gian lận tự động bằng thông tin hình ảnh trong phòng thi trắc nghiệm.

Việc áp dụng kiến thức liên ngành gồm IoT, Điện - Điện tử, Trí tuệ nhân tạo được sử dụng ngày càng nhiều nhằm nâng cao chất lượng cuộc sống nói chung [1-3] và trong quản lý giáo dục nói riêng [16-18]. Quản lý phòng thi thông minh cũng đã được một số nhóm nghiên cứu đề xuất, tuy nhiên chỉ là các hệ thống đơn lẻ chỉ là phần mềm phát hiện hành vi gian lận [18-20] hoặc bộ CSDL về hành vi gian lận trong phòng thi chưa được định nghĩa một cách chi tiết và tường minh, cũng như chưa công bố bộ CSDL cho cộng đồng nghiên cứu dùng chung hoặc mô hình phát hiện hành vi gian lận còn khá đơn giản [19, 20]. Như chúng ta thấy, mạng nơ ron tích chập ngày càng cho thấy nhiều ưu điểm và được áp dụng trong nhiều lĩnh vực khác nhau như xử lý ảnh [1], xử lý ngôn ngữ tự nhiên[4],... Trong xử lý ảnh, hiện có rất nhiều loại kiến trúc mạng nơ ron tích chập đã được các nhóm nghiên cứu đề xuất với nhiều mục đích khác nhau và chia thành nhiều khía cạnh khác nhau như phát hiện hay theo vết đối tượng trong ảnh hoặc chuỗi ảnh (video) [5], nhận dạng[12],... Mô hình YOLO là một trong những kiến trúc CNN khá hiệu quả trong bài toán phát hiện và nhận dạng đối tượng trong ảnh được công bố cho cộng đồng dùng chung. Mô hình này được công bố phiên bản đầu tiên năm 2016 [6]. Cho đến nay, mô hình YOLO đã có

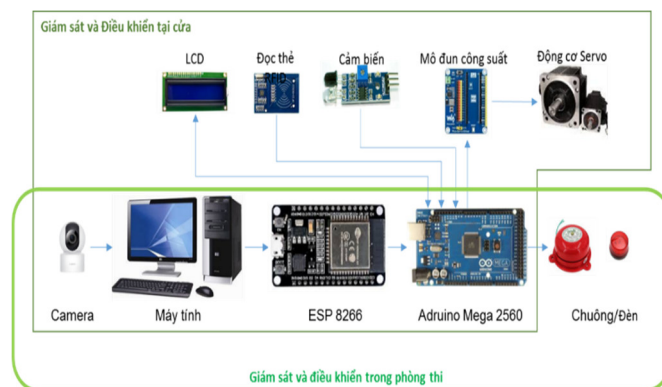
nhiều phiên bản khác nhau được công bố từ YOLO V1-V12. Trong các mô hình YOLO hiện có, mô hình YOLO V5 [7] được xem như một trong những mô hình có cấu trúc đơn giản hơn cả mà lại có hiệu quả khá tốt và số lượng tham số mô hình ít hơn các mô hình công bố sau đó [8-11] trong một số trường hợp đã chứng minh bằng thực nghiệm [1, 4]. Do đó, trong nghiên cứu này chúng tôi sẽ sử dụng mô hình YOLO V5 để thực hiện các đánh giá thử nghiệm.

Hệ thống đề xuất của chúng tôi có hai phần chính gồm: (a1) Mô hình phát hiện và nhận dạng khuôn mặt, trong đó, chúng tôi sẽ tiến hành đánh giá, so sánh giữa các mô hình nhận dạng khuôn mặt hiện nay như kiến trúc Dlib kết hợp với bộ phân lớp SVM trong [2], mô hình Facenet [3] và mô hình YOLO V5 [7]. Từ đánh giá thử nghiệm chúng tôi sẽ lựa chọn ra giải pháp phù hợp cho hệ thống kiểm tra tại cửa; (a2) Mô hình phát hiện hành vi gian lận, chúng tôi định nghĩa bộ CSDL hành vi gian lận trong phòng thi trắc nghiệm, sau đó tiến hành gán nhãn và đánh giá kết quả nhận dạng trên các mô hình YOLO V5. Trong nghiên cứu này, chúng tôi thực hiện đánh giá độ chính xác và thời gian đáp ứng phiên bản YOLO V5 với hàm mục tiêu cải tiến để có mô hình tốt hơn. Các thử nghiệm đánh giá trên bộ cơ sở dữ liệu (CSDL) hành vi gian lận trong phòng thi trắc nghiệm do chúng tôi tự định nghĩa, thu thập, gán nhãn và đặt tên bộ CSDL này là EPUExCheating.

Phần tiếp theo của bài báo gồm các phần sau đây: Phần 2 mô tả chi tiết giải pháp đề xuất. Kết quả thử nghiệm được trình bày trong Phần 3. Phần 4 là mục cuối cùng sẽ trình bày kết luận và hướng phát triển trong thời gian tiếp theo.

2. GIẢI PHÁP ĐỀ XUẤT

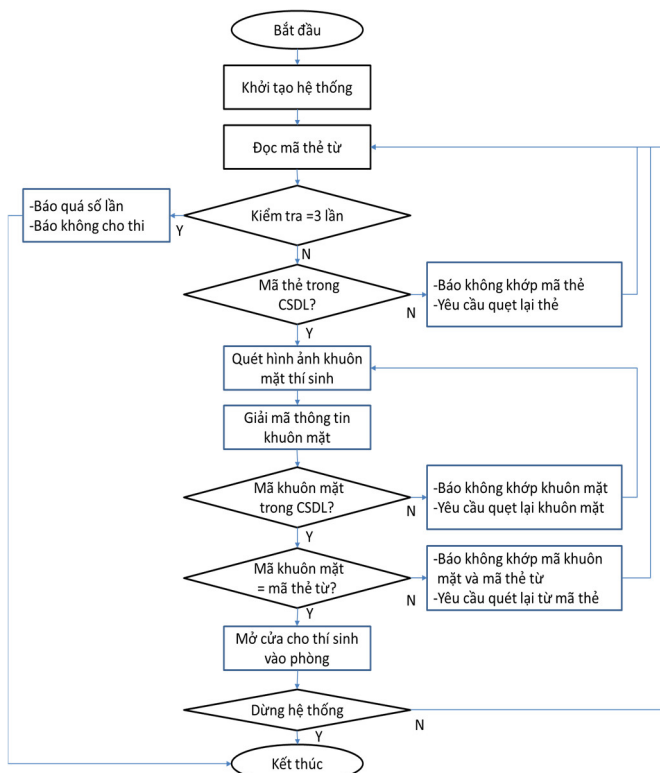
Hệ thống giám sát phòng thi gồm hai khối chính là khối giám sát tại cửa (khối đường bao màu xanh nhạt phía trên) và khối giám sát trong phòng thi (khối đường bao màu xanh lá cây đậm phía dưới) trong hình 1.



Hình 1. Sơ đồ khối hệ thống phát hiện gian lận trong phòng thi

Thành phần đầu tiên trong hệ thống là khối giám sát tại cửa. Trong đó, đầu vào của khối này gồm cảm biến phát hiện vật cản loại HC-SR04 để phát hiện có người. Đầu đọc thẻ RFID RC522 để đọc mã thẻ sinh viên. Camera sử dụng loại Camera HiK-Vision DS-2CD2643G2-IZS cung cấp ảnh màu RGB với tốc độ thu thập 30 (fps), độ phân giải ảnh 2688x1520 (pixels). Để thu thập dữ liệu hình ảnh, phát hiện và nhận dạng khuôn mặt của thí sinh khi vào cửa. Đầu ra gồm LCD dùng sử dụng loại LCD 1602 I2C hai dòng 16 ký tự để hiển thị thông báo chỉ dẫn cho thí sinh. Động cơ servo loại MG996 để mở cửa cho thí sinh vào phòng thi sau khi xác minh định danh thí sinh đúng. Thành phần tiếp theo của hệ thống phát hiện gian lận trong phòng thi là khối giám sát trong phòng thi. Khối này có thiết kế phần cứng gồm Camera, máy tính để phát hiện hành vi gian lận được kết nối trực tiếp với mô đun ESP8266 giao tiếp dữ liệu với mô đun điều khiển trung tâm là Arduino Mega 2560 và điều khiển chuông rung hoặc đèn cảnh báo. Ngoài ra, cảnh báo từ xa bằng cách báo tin trên ứng dụng cài đặt trên thiết bị thông minh như máy tính hoặc điện thoại di động.

2.1. Hệ thống kiểm tra ngay tại cửa



Hình 2. Lưu đồ thuật toán quá trình chống gian lận tại cửa phòng thi

Hệ thống nhận dạng khuôn mặt kết hợp với mã thẻ từ sẽ được sử dụng để chống gian lận theo hai lớp bảo mật như trong lưu đồ thuật toán thể hiện trong hình 2. Trong đó, hệ thống điều khiển ở cửa sẽ yêu cầu quét thẻ từ sau

đó kiểm tra mã thẻ từ đó có trong ca thi đó hay không. Nếu mã thẻ từ đúng thì yêu cầu thí sinh kiểm tra khuôn mặt. Nếu kết quả nhận dạng khuôn mặt của thí sinh không có trong CSDL ca thi thì sẽ yêu cầu quét lại khuôn mặt. Ngược lại, nếu có mã khuôn mặt thì hệ thống sẽ kiểm tra sự tương thích giữa mã khuôn mặt và mã thẻ từ. Nếu cả hai thông tin khuôn mặt và mã thẻ từ khớp nhau thì mở cửa và cho thí sinh được vào phòng thi.

Trong đó, phần mềm nhận dạng khuôn mặt sử dụng ba mô hình Dlib và SVM [2], mạng học sâu FaceNet [3] và YOLO V5 [7] để đánh giá quá trình nhận dạng khuôn mặt. Trong đó, chúng tôi tiến hành thu thập dữ liệu của 25 người tương đương với số lượng thí sinh trong một phòng thi. Trong đó, camera hướng đối diện với khuôn mặt của người thu. Mỗi người thu thập 100 ảnh và mỗi người thu trên 05 hướng khuôn mặt khác nhau mỗi hướng 20 ảnh gồm: hướng chính diện, xoay lệch trái, xoay lệch phải, xoay lệch trên, xoay lệch dưới.

Để đánh giá và lựa chọn mô hình chúng tôi chia dữ liệu của mỗi người thành $N = 10$ phần ngẫu nhiên và có số lượng bằng nhau. Giao thức để chia dữ liệu training, validation và testing được thực hiện tương tự như đã trình bày trong [14]. Kết quả độ chính xác của một lần thử nghiệm là Acc_i ($i = (1, \dots, N)$). Do đó, độ chính xác trung bình của mỗi mô hình trên toàn bộ CSDL được tính như công thức (1):

$$Acc(\%) = \frac{\sum_{i=1}^{N=10} Acc_i}{N} \quad (1)$$

2.2. Hệ thống phát hiện gian lận trong phòng thi

Hệ thống phát hiện hành vi gian lận trong phòng thi được đề xuất gồm các nội dung như sau: (1) Định nghĩa các hành động gian lận trong phòng thi; (2) Thu thập CSDL và gán nhãn; (3) Phát hiện và nhận dạng hành vi gian lận trong phòng thi; (4) Cảnh báo có hành vi gian lận, chi tiết được trình bày trong các phần dưới đây.

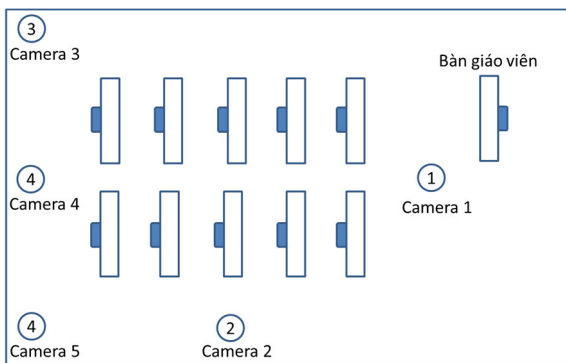
2.2.1. Định nghĩa, thu thập và gán nhãn bộ CSDL hành vi gian lận trong phòng thi

Trong nghiên cứu này, chúng tôi định nghĩa bốn hành vi gian lận cơ bản trong phòng thi gồm: sử dụng điện thoại, sử dụng tài liệu, sử dụng đồng hồ thông minh và trao đổi bài như trong bảng 1.

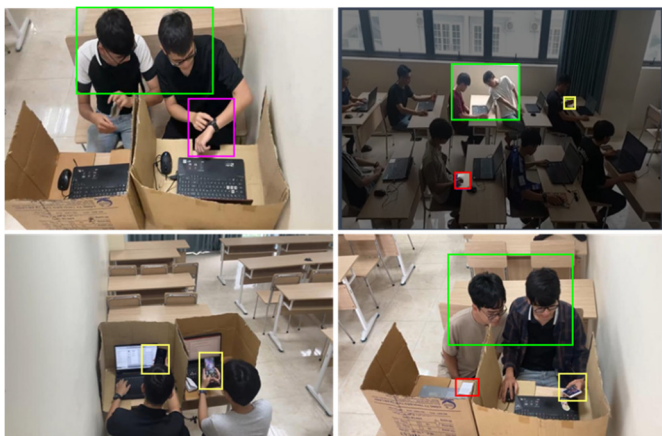
Bảng 1. Số lượng mẫu trong bộ CSDL EPUExCheating

Nhãn	Loại hành vi gian lận trong phòng thi	Số lượng ảnh
1	Sử dụng điện thoại	10230
2	Sử dụng tài liệu	9147
3	Sử dụng đồng hồ thông minh	10821
4	Trao đổi bài	12691

Môi trường để thu thập dữ liệu như thể hiện trong hình 3. Trong đó, phòng thi gồm một bàn giáo viên và 10 bàn thí sinh. Mỗi bàn sẽ chỉ có một hoặc hai người ngồi. Sử dụng từ 1 đến 5 camera để thu thập dữ liệu. Các camera được treo cách mặt sàn 2m và chếch hướng về chỗ thí sinh ngồi. Phòng thi mô phỏng có sức chứa tối đa 10 thí sinh và các thí sinh có thể thi xong tại các thời điểm khác nhau và khi thi xong thì thí sinh có thể được phép rời phòng nên sẽ có số lượng thí sinh trong phòng thi là không cố định. Do đó, chúng tôi thiết kế môi trường phòng thi với số lượng người trong phòng dao động từ 1 đến 10 thí sinh.



Hình 3. Minh họa môi trường thu cơ sở dữ liệu EPUExCheating



Hình 4. Minh họa một số hình ảnh trong bộ CSDL EPUExCheating

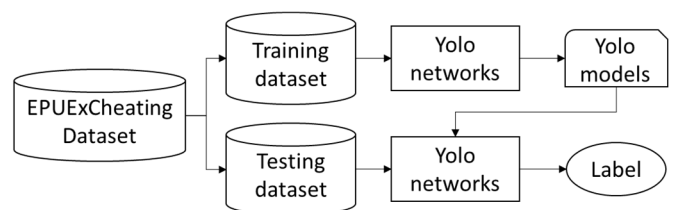
Hình 3 minh họa một số hình ảnh thu được trong phòng thi trắc nghiệm trên máy tính. Chúng tôi đặt tên bộ CSDL này là EPUExCheating. Hình 4 minh họa các hình ảnh thu được lần lượt từ các camera khác nhau trong bộ CSDL này của chúng tôi. Trong đó, hình bao màu xanh lá cây là các nhãn trao đổi bài, màu đỏ là các nhãn sử dụng tài liệu, màu vàng là các nhãn sử dụng điện thoại và màu tím là các nhãn sử dụng đồng hồ.

Tổng số video chúng tôi thu được là 50 video, mỗi video có độ dài trung bình là 1 đến 10 phút. Sau khi thu dữ liệu, chúng tôi tách lấy mẫu 10 khung hình lấy một lần và thực hiện gán nhãn bằng công cụ Roboflow. Toàn bộ

CSDL đã gán nhãn được thống kê theo từng nhãn như trong bảng 1. Bộ CSDL này được chúng tôi sử dụng để đánh giá mô hình phát hiện vi gian lận trong phòng thi bằng hình ảnh sẽ trình bày chi tiết trong mục 2.3.2. Toàn bộ quá trình đánh giá định lượng được thực hiện chi tiết trong mục 3. Từ kết quả đánh giá, chúng tôi lựa chọn mô hình và sử dụng toàn bộ CSDL EPUExCheating này để huấn luyện mô hình thực tế.

2.2.2. Phần mềm phát hiện gian lận trong phòng thi

Trong phần phát hiện và nhận dạng hành vi gian lận trong phòng thi sử dụng thông tin hình ảnh, chúng tôi thực hiện đánh giá trên phiên bản YOLO V5 [7]. Trong đó quá trình huấn luyện và đánh giá mô hình được thực hiện như minh họa trong hình 5.



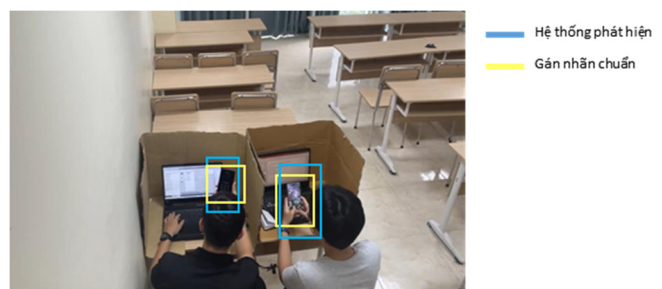
Hình 5. Quá trình huấn luyện và thử nghiệm mô hình phát hiện gian lận trong phòng thi

Hàm mất mát của mô hình YOLO V5 [7] sử dụng là hàm Classification Loss, Confidence Loss và hàm CIoU Loss như biểu diễn trong công thức (2) sau đây:

$$L_{YOLOV5} = L_{Classification} + L_{Confidence} + L_{CIoU} \quad (2)$$

Trong nghiên cứu này, chúng tôi sẽ thực hiện điều chỉnh làm mất mát cho mô hình YOLO V5 bằng cách kết hợp hàm mất mát gốc của mô hình YOLO V5 trong công thức (2) với hàm Focal Loss (FC) [15] như trình bày trong công thức (3) sau đây:

$$L = (1 - \alpha)(L_{Classification} + L_{Confidence} + L_{CIoU}) + \alpha L_{FC} \quad (3)$$



Hình 6. Minh họa dữ liệu phát hiện và dữ liệu được gán nhãn với nhãn sử dụng điện thoại

Trong đó, hệ số α được sử dụng với các giá trị khác nhau để điều chỉnh mức độ tham gia của hàm mất mát tiêu chuẩn của mô hình YOLO V5 với thành phần Focal Loss. Nếu giá trị α càng nhỏ thì trọng số của hàm mất mát

gốc của mô hình YOLO V5 lớn, trong khi trọng số ảnh hưởng của thành phần Focal Loss nhỏ và ngược lại. Để đánh giá hiệu quả của giải pháp đề xuất trong công thức (3) cũng như đánh giá sự ảnh hưởng của thành phần Focal Loss lên hiệu quả của mô hình YOLO V5 sẽ được chúng tôi trình bày chi tiết trong mục 3. Để đánh giá hệ thống phát hiện hành vi gian lận trong phòng thi, chúng tôi sử dụng giao thức thử nghiệm "leave-one-subject-out" [1] để chia CSDL cho tập train, validation và test. Mỗi lần thử nghiệm sẽ đánh giá các tiêu chí gồm độ chính xác P (Precision), độ triệu hồi R (Recall) và độ chính xác trung bình mAP (Mean Average Precision). Trong đó, nếu vùng gán nhãn chuẩn là màu vàng, vùng phát hiện từ mô hình là màu xanh dương như minh họa trong hình 6. Khi đó, công thức xác định độ chính xác được giải thích chi tiết như sau:

Độ chính xác P được định nghĩa là tỉ lệ của vùng phát hiện đúng từ mô hình (giao của vùng màu xanh dương và màu vàng) chia cho vùng dữ liệu được gán nhãn - Ground Truth (màu vàng). Công thức tính độ chính xác Precision của mỗi vùng ảnh hành vi gian lận được minh họa bởi công thức (4):

$$P = \frac{TP}{TP+FP} \quad (4)$$

Độ triệu hồi R là tỉ lệ của vùng phát hiện đúng từ mô hình (giao của vùng màu xanh dương và màu vàng) chia cho vùng phát hiện từ mô hình (màu xanh dương). Công thức tính độ triệu hồi Recall được xác định bởi công thức (5):

$$R = \frac{TP}{TP+FN} \quad (5)$$

Độ chính xác trung bình mAP của tất cả các lớp dữ liệu (N = 4 lớp) được định nghĩa theo công thức (6):

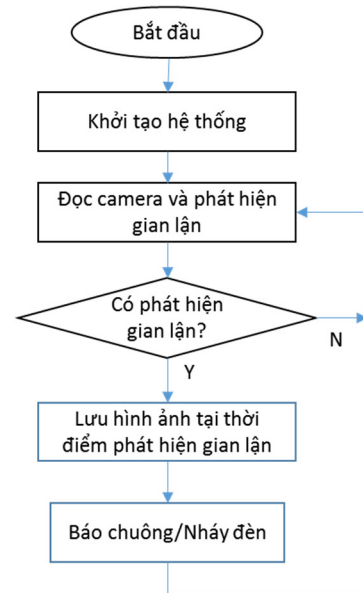
$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

Trong đó, TP (True Positive) là tổng số vùng ảnh dự đoán từ mô hình là đúng được phân lớp vào đúng vùng được gán nhãn; FP (False Positive) tổng số vùng ảnh mô hình dự đoán sai được phân lớp vào vùng được gán nhãn đúng; FN (False Negative) tổng số vùng ảnh mô hình dự đoán sai được phân lớp vào vùng gán nhãn sai; và TN (True Negative) là tổng số vùng ảnh mô hình dự đoán đúng được phân lớp vào vùng gán nhãn sai, N là tổng số lớp được gán nhãn. Kết quả đánh giá hệ thống sẽ được thực hiện và trình bày chi tiết trong mục 3.

2.2.3. Phát hiện và cảnh báo hành vi gian lận

Sau khi phát hiện có hành vi gian lận, hệ thống sẽ chụp lại ảnh nghi ngờ gian lận và lưu vào ổ cứng của máy tính. Ngoài ra, hệ thống còn phát chuông rung hoặc báo đèn

cho giám thị biết trong phòng thi có hành vi gian lận để giám thị có hình thức kiểm tra lại thông tin hình ảnh đã được lưu trên ổ cứng máy tính. Trên cơ sở đó, giám thị sẽ có hình thức xử lý hành vi gian lận trong phòng thi một cách phù hợp. Hình 7 minh họa lưu đồ thuật toán hoạt động của hệ thống phát hiện hành vi gian lận trong phòng thi trắc nghiệm.



Hình 7. Lưu đồ thuật toán quá trình chống gian lận bên trong phòng thi

3. KẾT QUẢ THỬ NGHIỆM

Chúng tôi thực hiện các thử nghiệm trên máy tính máy sử dụng CPU Intel Core i5-11400H, GPU NVIDIA GeForce GTX 1650, bộ nhớ Ram 8GB. Ngôn ngữ lập trình sử dụng là ngôn ngữ Python. Các thử nghiệm được thực hiện trên CSDL EPUExCheating như mô tả trong mục 2.3.1. Trong đó, các thử nghiệm cài đặt với kích thước mẻ (batch size) là 64; tốc độ huấn luyện mô hình (learning rate) thực hiện chế độ tự dừng sớm tự động (early stopping) với giá trị dao động trong dải từ 10^{-6} đến 10^{-4} ; số lần lặp để huấn luyện mô hình (epochs) là 100. Các thử nghiệm thực hiện sử dụng các mạng gồm YOLO V5 [7], FaceNet [3], Dlib và SVM [2].

Các thử nghiệm được tiến hành trong nghiên cứu này gồm: (1) Đánh giá độ chính xác của các mô hình nhận dạng khuôn mặt; (2) Đánh giá độ chính xác phát hiện hành vi gian lận; (3) Đánh giá kích thước mô hình và thời gian đáp ứng.

3.1. Độ chính xác nhận dạng khuôn mặt

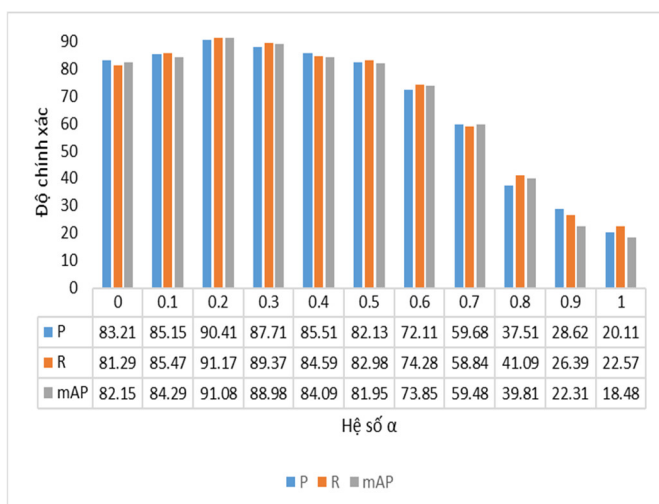
Trong phần 2.1 đã trình bày thu thập CSDL khuôn mặt, cài đặt thử nghiệm cho đánh giá mô hình nhận dạng khuôn mặt. Kết quả thử nghiệm gồm độ chính xác và thời gian đáp ứng như trình bày trong bảng 2.

Bảng 2. Độ chính xác phát hiện khuôn mặt

Phương pháp	Độ chính xác (%)	Thời gian đáp ứng trên GPU (FPS)	Thời gian đáp ứng trên CPU (FPS)
Dlib + SVM [2]	80,56	7	5
YOLO V5 [7]	85,17	25	16
FaceNet [3]	98,91	22	13

Bảng 2 cho thấy phương pháp Dlib kết hợp với bộ phân lớp SVM trong [2] đạt độ chính xác thấp nhất trên bộ CSDL khuôn mặt của chúng tôi chỉ với 80,56% trong tốc độ khá chậm với 7fps trên CPU và 5fps trên GPU. Phương pháp FaceNet [3] có thời gian đáp ứng chậm hơn không đáng kể so với YOLO V5 (22 fps so với 25 fps trên GPU và 13fps so với 16fps trên CPU) trong khi độ chính xác của FaceNet [3] lại có độ chính xác cao nhất với 98,91% và cao hơn nhiều so với hai phương pháp còn lại tại 85,17% của YOLO V5 và 80,56% của Dlib + SVM [2]. Do đó, chúng tôi sử dụng FaceNet [3] để triển khai mô hình kiểm tra và định danh tại cửa phòng thi.

3.2. Độ chính xác phát hiện hành vi gian lận



Hình 8. Độ chính xác phát hiện hành vi gian lận trong phòng thi trắc nghiệm trên bộ CSDL EPUExCheating với các giá trị hệ số α khác nhau

Trong phần này chúng tôi sẽ thực hiện đánh giá độ chính xác Pre, độ triệu hồi Recall và độ chính xác trung bình tại IoU = 0,5 (mAP) của mô hình YOLO V5 [7]. Trong đó, hệ số α trong công thức (3) được thực hiện tại các giá trị khác nhau để xem xét sự ảnh hưởng của giá trị hàm mất mát cải tiến so với hàm mất mát ban đầu của mô hình YOLO V5. Thử nghiệm được đánh giá trên bộ CSDL EPUExCheating. Kết quả được biểu diễn như trong hình 8.

Kết quả trong hình 8 cho thấy:

- Mô hình YOLO V5 với hàm mục tiêu gốc trong [7] tương ứng công thức (2) hoặc tương ứng công thức (3) tại

giá trị $\alpha = 0$ có độ chính xác trên cả ba tiêu chí Pre, Recall và @mAP50 lần lượt tại 83,21%; 81,29% và 82,15% có giá trị cao hơn khi chỉ sử dụng hàm Focal Loss như trong công thức (3) tại giá trị $\alpha = 1$ chỉ đạt các giá trị lần lượt tại 20,11%; 22,57% và 18,48%.

- Mô hình đạt độ chính xác cao nhất tại $\alpha = 0,2$ tương ứng với các giá trị 90,41%; 91,17% và 91,08%. Điều này cho thấy việc áp dụng thành phần Focal Loss vào hàm mục tiêu giúp mô hình YOLO V5 cải thiện khá đáng kể độ chính xác trên bộ CSDL EPUExCheating trên cả ba tiêu chí độ chính xác. Tuy nhiên, khi tăng dần trọng số α để tăng mức ảnh hưởng của giá trị hàm Focal Loss thì các giá trị độ chính xác Pre, Recall và mAP cùng có xu hướng giảm dần.

4. THẢO LUẬN VÀ KẾT LUẬN

Bài báo đã thiết kế giải pháp toàn diện cho hệ thống quản lý phòng thi. Hệ thống đề xuất của chúng tôi giám sát tự động ngay từ cửa thông qua thẻ từ và nhận dạng khuôn mặt bằng hình ảnh. Ngoài ra, trong phòng thi còn có thể tự động phát hiện và cảnh báo hành vi gian lận một cách tự động, hỗ trợ hữu ích cho giám thị khi quản lý phòng thi. Chúng tôi đã định nghĩa, thu thập và gán nhãn được một bộ CSDL về hành vi gian lận trong phòng thi trắc nghiệm. Ngoài ra, nghiên cứu cũng đánh giá định lượng hệ thống phát hiện hành vi gian lận trong phòng thi trên mô hình YOLO V5 với hàm mục tiêu đề xuất mới. Kết quả cho thấy độ chính xác mô hình cải thiện đáng kể so với mô hình gốc lên tới 7,2% trên tiêu chí Precision, 9,88% trên tiêu chí Recall và 8,93% trên độ chính xác mAP. Trong thời gian sắp tới, chúng tôi sẽ tiếp tục thu thập thêm dữ liệu hành vi bất thường trong phòng thi và cải thiện hiệu quả của hệ thống phát hiện và nhận dạng hành vi gian lận trong phòng thi trắc nghiệm. Trên cơ sở đó, chúng tôi sẽ tiến hành tạo CSDL người dùng và liên kết với hệ thống hiện có để tạo ứng dụng hoàn thiện. Sau đó, chúng tôi sẽ thực hiện thêm các thử nghiệm hệ thống trên các mô hình mạng khác nhau để so sánh và có thêm cơ sở giúp lựa chọn giải pháp. Ngoài ra, chúng tôi sẽ tiến hành tích hợp hệ thống để có mô hình nhỏ gọn hơn, dễ dàng lắp đặt và triển khai trong thực tế.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi quá trình thực hiện đề tài nghiên cứu khoa học cấp trường Đại học Điện Lực năm 2025 dành cho viên chức và người lao động với tiêu đề: **"Nghiên cứu, thiết kế hệ thống giám sát và quản lý trong phòng thi trắc nghiệm trên máy tính"**.

TÀI LIỆU THAM KHẢO

- [1]. Hồ Anh Dũng, Đoàn Thị Hương, Giang, Trần Đình Hùng, Ma Khánh Tùng, Nguyễn Huyền Tiến An, Bùi Thị Duyên, "Hệ thống phát hiện khói và cháy đa thông minh đa thể thức," *Tạp chí Nghiên cứu Khoa học và Công nghệ quân sự*, 97, 2024. doi: 10.54939/1859-1043.j.mst.97.2024.138-147, 08-2024.
- [2]. Huong-Giang Doan, Ngoc-Trung Nguyen, "Realtime Face Recognition for Intelligent Door System," *International Journal of Emerging Trends in Engineering Research (IJETER)*, 8, 9, 5226-5232, 2020.
- [3]. F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 815-823, 2015.
- [4]. K. S. Varshitha, C. G. Kumari, M. Hasvitha, S. Fiza A. K, V. Rachapudi, "Natural Language Processing using Convolutional Neural Network," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 362-367, 2023. doi: 10.1109/ICCMC56507.2023.10083608.
- [5]. Dinh-Ha Nguyen, Trung-Hieu Le, Thanh-Hai Tran, Hai Vu, Thi-Lan Le, Huong-Giang Doan, "Hand segmentation under different viewpoints by combination of Mask R-CNN with tracking," in *2018 5th Asian Conference on Defense Technology (ACDT)*, 14-20, 2018.
- [6]. Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali, "YOLOv1: You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788, 2016. doi: 10.1109/CVPR.2016.91.
- [7]. Glenn Jocher, *Yolov5 in pytorch*. <https://github.com/ultralytics/yolov5>, 06 2020.
- [8]. Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, Xiaolin Wei, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," *Computer Vision and Pattern Recognition*, 2022.
- [9]. Wang Chien-Yao, Bochkovskiy Alexey, Liao Hong-yuan, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [10]. Jocher G., Chaurasia A., Qiu J., *YOLO by Ultralytics*. GitHub. 1 January 2023. Available online: <https://github.com/ultralytics/ultralytics>.
- [11]. Wang C., Yeh I., Liao H., "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," in *Lecture Notes in Computer Science*, 15089, 2024.
- [12]. Huong-Giang Doan, Ngoc-Trung Nguyen, "End-to-end multiple modals deep learning system for hand posture recognition," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, 27, 1, 214-221, 2022.
- [13]. M. A. Turk, A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, HI, USA, 586-591, 1991. doi: 10.1109/CVPR.1991.139758.
- [14]. Huong-Giang Doan, Anh-Dung Ho, Trung-Hieu Nguyen, Ngoc-Trung Nguyen, "Improvement of performance of human detection in abnormal crowd using knowledge distillation for YOLO network," *HaUI Journal of Science and Technology*, 60, 4, 39-44, 2024. doi: <http://doi.org/10.57001/huih5804.2024.124>.
- [15]. T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 2, 318-327, 2020.
- [16]. Thi-Oanh Ha, Hoang-Nhat Tran, Hong-Quan Nguyen, Thanh-Hai Tran, Phuong-Dung Nguyen, Huong-Giang Doan, Van-Toi Nguyen, Hai Vu, Thi-Lan Le, "Improvement of People Counting by Pairing Head and Face Detections from Still Images," in *2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 2021.
- [17]. Thu-Hien Le, Hoang-Nhat Tran, Phuong-Dung Nguyen, Hong-Quan Nguyen, Thuy-Binh Nguyen, Thanh-Hai Tran, Hai Vu, Thi-Thao Tran, Thi-Lan Le, "Spatial and Temporal Hand-Raising Recognition from Classroom Videos Using Locality, Relative Position-Aware Non-Local Networks and Hand Tracking," *Vietnam Journal of Computer Science*, 02, 10, 243-271, 2023.
- [18]. Razan Bawarith, Abdullah Basuhail, Anas Fattouh, Shehab Gamalel-Din, "E-exam Cheating Detection System," *International Journal of Advanced Computer Science and Applications*, 8, 4, 2017.
- [19]. <https://hust.edu.vn/vi/news/tin-tuc-su-kien/phan-mem-chong-gian-lan-thi-cu-duoc-nhom-sinh-vien-thu-nghiem-thanh-cong-648900.html>
- [20]. <https://www.hauai.edu.vn/vn/page/ts/detail/63144>

AUTHORS INFORMATION**Doan Thi Huong Giang¹, Ho Doan Bao Chau²**¹Faculty of Control and Automation, Electric Power University, Vietnam²School of Information and Communications Technology, Hanoi University of Science and Technology, Vietnam