

# Phát Hiện Hành Vi Gian Lận Trong Thi Cử

Trần Ngọc Duyên, Phạm Thị Huyền Trang, Nguyễn Trọng Anh, Nguyễn Anh Dũng, Trương Văn An  
Nhóm 9, Khoa Công Nghệ Thông Tin  
Trường Đại Học Đại Nam, Việt Nam

**ThS. Nguyễn Văn Nhân, ThS. Lê Trung Hiếu**  
Giảng viên hướng dẫn, Khoa Công Nghệ Thông Tin  
Trường Đại Học Đại Nam, Việt Nam

*Tóm tắt nội dung*—Gian lận trong thi cử đã trở thành một vấn đề phổ biến và ngày càng phức tạp trong môi trường giáo dục hiện đại. Các hình thức gian lận truyền thống như quay cốp, trao đổi bài hay sử dụng tài liệu không được phép vẫn tồn tại, nhưng với sự phát triển của công nghệ, những phương thức gian lận tinh vi hơn đang được áp dụng. Đặc biệt, việc sử dụng các thiết bị công nghệ cao như điện thoại thông minh, tai nghe không dây, hay camera giấu kín đã làm cho việc giám sát thi cử trở nên khó khăn hơn đối với các giám thị.

Trong bối cảnh này, việc ứng dụng công nghệ trí tuệ nhân tạo (AI) và thị giác máy tính (Computer Vision) để tự động phát hiện và cảnh báo hành vi gian lận là một hướng đi tiềm năng nhằm nâng cao hiệu quả giám sát và đảm bảo tính công bằng trong thi cử. Đề tài này tập trung vào việc xây dựng hệ thống phát hiện gian lận thi cử tự động sử dụng mô hình YOLOv11 – một trong những mô hình nhận diện vật thể tiên tiến nhất hiện nay. Hệ thống này có khả năng giám sát liên tục thông qua các camera được lắp đặt trong phòng thi, tự động nhận diện và phân loại các hành vi gian lận như trao đổi bài, nhìn bài người khác, hoặc sử dụng tài liệu và điện thoại trái phép.

Bên cạnh đó, hệ thống còn có chức năng cảnh báo thời gian thực thông qua giao diện web, giúp giám thị dễ dàng theo dõi và can thiệp kịp thời. Những kết quả đạt được từ mô hình không chỉ nâng cao khả năng phát hiện chính xác hành vi gian lận, mà còn giảm tải áp lực cho giám thị trong các kỳ thi đông đúc và quy mô lớn.

## I. GIỚI THIỆU VỀ YOLOV11

YOLO (You Only Look Once) là một trong những mô hình nổi tiếng và hiệu quả nhất trong lĩnh vực phát hiện đối tượng (object detection). Được giới thiệu lần đầu vào năm 2016, YOLO đã không ngừng phát triển qua nhiều phiên bản, và đến phiên bản mới nhất YOLOv11, mô hình đã đạt được nhiều cải tiến vượt trội về độ chính xác và tốc độ. YOLOv11 là phiên bản cải tiến của các mô hình YOLO trước đó, kết hợp các kỹ thuật hiện đại như tối ưu hóa mạng học sâu, sử dụng các kỹ thuật huấn luyện dữ liệu tiên tiến và kiến trúc mới nhằm tối ưu khả năng nhận diện đối tượng trong thời gian thực. Một trong những đặc điểm nổi bật của YOLOv11 là khả năng xử lý nhanh chóng với độ trễ thấp, đồng thời vẫn giữ được độ chính xác cao, giúp nó trở thành lựa chọn hàng đầu cho các ứng dụng yêu cầu giám sát thời gian thực như phát hiện gian lận trong thi cử.

## II. LÝ DO LỰA CHỌN YOLOV11

### A. Tốc độ xử lý nhanh

YOLOv11 có khả năng xử lý hình ảnh và video trong thời gian thực với tốc độ rất nhanh. Điều này đặc biệt quan trọng



Hình 1. Hình ảnh logo YOLOv11

trong việc giám sát phòng thi, nơi hệ thống cần phát hiện và cảnh báo ngay lập tức các hành vi gian lận để giám thị can thiệp kịp thời.

### B. Độ chính xác cao

YOLOv11 đạt được độ chính xác rất cao trong việc nhận diện và phân loại đối tượng. Điều này giúp mô hình phát hiện chính xác các hành vi gian lận, từ việc thí sinh nhìn bài của người khác đến việc sử dụng các thiết bị như điện thoại hay tài liệu trái phép.

### C. Khả năng nhận diện đa đối tượng

Một trong những điểm mạnh của YOLOv11 là khả năng nhận diện nhiều đối tượng khác nhau trên cùng một khung hình. Trong một phòng thi với nhiều thí sinh, mô hình có thể phát hiện đồng thời nhiều hành vi gian lận diễn ra ở các vị trí khác nhau.

### D. Tối ưu hóa tài nguyên

Mặc dù YOLOv11 có hiệu suất cao, nhưng nó vẫn hoạt động tốt trên các hệ thống phần cứng không quá mạnh. Điều này giúp hệ thống giám sát có thể triển khai trên nhiều thiết bị giám sát phổ thông mà không cần phải đầu tư vào các thiết bị có cấu hình quá cao.

### E. Tính ứng dụng cao trong giám sát thời gian thực

Với yêu cầu giám sát liên tục và phát hiện hành vi gian lận ngay khi nó xảy ra, YOLOv11 là lựa chọn lý tưởng. Khả năng phát hiện đối tượng nhanh và chính xác của mô hình giúp đảm

bảo hệ thống giám sát có thể hoạt động ổn định và hiệu quả trong suốt thời gian diễn ra kỳ thi.

### III. ĐẶC ĐIỂM CỦA YOLOv11

YOLOv11 (You Only Look Once, phiên bản 11) là một trong những mô hình tiên tiến nhất trong lĩnh vực phát hiện đối tượng. Mô hình này kế thừa và phát triển từ các phiên bản YOLO trước, với nhiều cải tiến đáng kể về cả độ chính xác và tốc độ xử lý. Dưới đây là những đặc điểm chính của YOLOv11:

#### A. Kiến trúc một lần duy nhất

YOLOv11 tiếp tục sử dụng triết lý "You Only Look Once", nghĩa là mô hình chỉ cần duyệt qua hình ảnh một lần duy nhất để phát hiện tất cả các đối tượng có trong đó. Điều này giúp mô hình hoạt động với tốc độ rất nhanh, phù hợp cho các ứng dụng thời gian thực.

#### B. Tăng cường độ sâu của mạng

YOLOv11 sử dụng mạng học sâu với nhiều lớp hơn, giúp nâng cao khả năng nhận diện đối tượng ở các kích thước và hình dạng khác nhau. Các lớp mạng sâu hơn giúp mô hình phát hiện các đối tượng nhỏ và tinh vi, điều rất quan trọng trong việc phát hiện hành vi gian lận trong thi cử.

#### C. Neck mới với CSPNet và PANet

YOLOv11 tích hợp CSPNet (Cross Stage Partial Network) và PANet (Path Aggregation Network) trong kiến trúc mạng của mình. Các kỹ thuật này giúp cải thiện sự kết hợp thông tin giữa các tầng mạng, làm tăng hiệu quả của quá trình nhận diện đối tượng.

#### D. Cải tiến quá trình huấn luyện

YOLOv11 sử dụng nhiều kỹ thuật tiên tiến như Data Augmentation (tăng cường dữ liệu), Hyperparameter Optimization (tối ưu siêu tham số), và Label Smoothing (làm mượt nhãn) để cải thiện hiệu suất. Những kỹ thuật này giúp mô hình có khả năng học tốt hơn, đồng thời giảm thiểu hiện tượng quá khớp (overfitting).

#### E. Hiệu suất cao trong thời gian thực

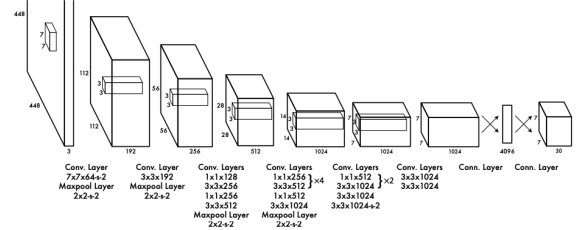
YOLOv11 không chỉ có khả năng nhận diện đối tượng chính xác mà còn có thể thực hiện nhanh chóng trong thời gian thực, ngay cả trên những thiết bị phần cứng có cấu hình trung bình. Điều này giúp cho YOLOv11 trở thành một giải pháp lý tưởng cho các hệ thống giám sát tự động, như giám sát thi cử.

#### F. Khả năng nhận diện nhiều đối tượng

YOLOv11 có thể phát hiện nhiều đối tượng khác nhau trên cùng một khung hình, điều này rất hữu ích trong các tình huống phức tạp như phòng thi, nơi có nhiều thí sinh và nhiều hành vi gian lận có thể diễn ra cùng lúc.

### IV. KIẾN TRÚC YOLOv11

YOLOv11 (You Only Look Once, phiên bản 11) tiếp tục phát triển dựa trên các nguyên tắc cốt lõi của các phiên bản trước, nhưng được cải tiến với nhiều yếu tố mới giúp tăng cường hiệu suất và độ chính xác trong phát hiện đối tượng. Kiến trúc của YOLOv11 được chia thành ba phần chính: Backbone, Neck, và Head.



Hình 2. Kiến trúc mô hình YOLOv11

#### A. Backbone

Backbone là phần đầu tiên trong mô hình YOLOv11, chịu trách nhiệm trích xuất các đặc trưng (features) từ hình ảnh đầu vào. YOLOv11 sử dụng một mạng học sâu có nhiều lớp chập (convolutional layers) nhằm phát hiện các đặc trưng quan trọng trong các vùng khác nhau của ảnh. Những đặc trưng này bao gồm kích thước, hình dạng, và màu sắc của các đối tượng.

YOLOv11 sử dụng kiến trúc CSPNet (Cross Stage Partial Network) làm backbone, giúp chia nhỏ quá trình trích xuất đặc trưng và giảm thiểu chi phí tính toán, đồng thời giữ nguyên được độ chính xác cao. CSPNet giúp tối ưu hóa quá trình học, cho phép mô hình học được nhiều đặc trưng đa dạng hơn mà không làm tăng số lượng tham số.

#### B. Neck

Phần Neck trong YOLOv11 có vai trò tổng hợp thông tin từ nhiều mức độ khác nhau của các đặc trưng được trích xuất từ backbone. Điều này giúp mô hình có thể nhận diện đối tượng ở nhiều kích thước và mức độ chi tiết khác nhau, từ các đối tượng nhỏ đến các đối tượng lớn.

YOLOv11 sử dụng PANet (Path Aggregation Network) trong Neck để kết hợp các đặc trưng từ các tầng khác nhau của backbone. PANet giúp cải thiện quá trình tổng hợp và truyền thông tin từ các đặc trưng thấp đến các đặc trưng cao, từ đó cải thiện khả năng phát hiện đối tượng của mô hình.

#### C. Head

Head là phần cuối cùng của mô hình YOLOv11, nơi các dự đoán về vị trí (bounding box), lớp đối tượng (object class), và xác suất được thực hiện. Từ các đặc trưng đã được tổng hợp bởi Neck, Head của YOLOv11 sẽ đưa ra dự đoán về những gì mô hình đã nhận diện trong hình ảnh.

YOLOv11 cải tiến quá trình dự đoán bằng cách sử dụng các cơ chế học sâu hiện đại như *GIoU Loss* (Generalized Intersection over Union) và *CIoU Loss* (Complete IoU), giúp cải thiện độ chính xác của các hộp chứa (bounding box) và đảm bảo rằng chúng khớp chính xác hơn với đối tượng thật trong hình ảnh.

#### D. Quá trình xử lý

YOLOv11 sử dụng một quá trình duy nhất để xử lý và phát hiện đối tượng trong hình ảnh. Khi nhận một hình ảnh đầu vào, mô hình chia nhỏ hình ảnh thành các ô lưới. Mỗi ô lưới sẽ chịu

trách nhiệm phát hiện một phần của đối tượng nếu nó nằm trong ô đó. Sau đó, mô hình dự đoán hộp chứa (bounding box), lớp của đối tượng, và xác suất đối tượng xuất hiện tại mỗi ô lưới.

Sự kết hợp giữa các thành phần Backbone, Neck, và Head cùng với khả năng xử lý thời gian thực đã giúp YOLOv11 trở thành một trong những mô hình phát hiện đối tượng hàng đầu, phù hợp với các ứng dụng yêu cầu giám sát liên tục và tốc độ cao như phát hiện gian lận trong thi cử.

## V. SO SÁNH YOLOV11 SO VỚI CÁC MÔ HÌNH KHÁC

Bảng I

SO SÁNH YOLOV11 VỚI CÁC MÔ HÌNH PHÁT HIỆN ĐỐI TƯỢNG KHÁC

Mô hình	Tốc độ (FPS)	Độ chính xác (mAP)	Thời gian thực	Yêu cầu tài nguyên
YOLOv11	60-70	97%	Có	Trung bình
YOLOv4	50-60	92%	Có	Trung bình
Faster R-CNN	7-10	89%	Không	Cao
SSD	22-30	85%	Có	Trung bình

## VI. PHƯƠNG PHÁP THỰC HIỆN

### A. Xây dựng và thu thập dữ liệu

Trong dự án này, chúng tôi sử dụng bộ dữ liệu SCB-dataset, kết hợp với dữ liệu tự thu thập từ phòng thi. Việc tiền xử lý dữ liệu bao gồm các bước: gán nhãn cho dữ liệu thu thập được, chuẩn hóa hình ảnh về cùng kích thước và định dạng, đồng thời sử dụng các kỹ thuật tăng cường dữ liệu để mở rộng tập dữ liệu huấn luyện, đảm bảo mô hình có khả năng tổng quát cao hơn.

### B. Huấn luyện mô hình YOLOv11

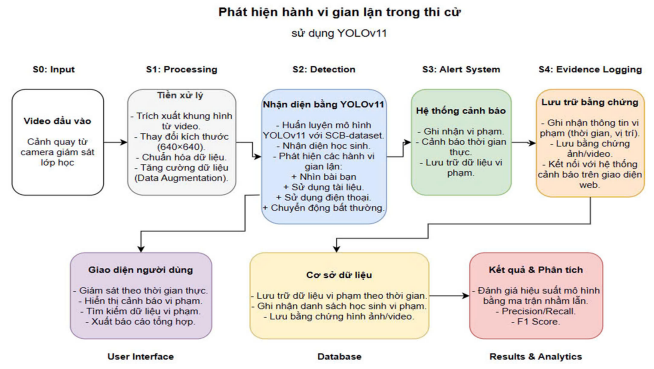
Mô hình YOLOv11 được huấn luyện nhằm tối ưu khả năng phát hiện chính xác các hành vi gian lận trong phòng thi. Chúng tôi thực hiện việc tinh chỉnh mô hình thông qua các tham số huấn luyện như learning rate, số epoch, và các kỹ thuật như Early Stopping để tránh overfitting. Hiệu suất của YOLOv11 được so sánh với các mô hình khác như Faster R-CNN và SSD để kiểm tra khả năng phát hiện chính xác và tốc độ xử lý.

### C. Phát triển hệ thống giám sát

YOLOv11 sau khi được huấn luyện sẽ được tích hợp vào hệ thống giám sát thời gian thực. Hệ thống này bao gồm việc xây dựng giao diện web giúp hiển thị cảnh báo khi phát hiện gian lận. Ngoài ra, chúng tôi còn tích hợp hệ thống cảnh báo tự động qua các kênh như email, Telegram hoặc loa báo động để cảnh báo ngay lập tức cho người giám sát.

## VII. SƠ ĐỒ HỆ THỐNG

Hình dưới đây minh họa kiến trúc hệ thống phát hiện gian lận trong thi cử. Dữ liệu từ camera giám sát được truyền về hệ thống máy chủ để xử lý và phát hiện gian lận. Hệ thống này hoạt động liên tục để giám sát thời gian thực và phát hiện hành vi gian lận, sau đó gửi cảnh báo đến người giám sát thông qua các kênh liên lạc đã tích hợp.



Hình 3. Kiến trúc mô hình YOLOv11 được sử dụng trong bài

## VIII. CÁC THIẾT BỊ PHẦN CỨNG

Hệ thống sử dụng các camera giám sát được đặt trên bàn hoặc gắn trên cao để có thể ghi lại hình ảnh hoặc video của thí sinh trong quá trình thi. Hình ảnh và video được truyền về hệ thống máy chủ hoặc cloud để xử lý. Hệ thống máy chủ cần có cấu hình đủ mạnh để đảm bảo việc xử lý thời gian thực và lưu trữ dữ liệu liên tục.

## IX. KẾT QUẢ THỰC NGHIỆM

### A. Độ lỗi trong quá trình huấn luyện

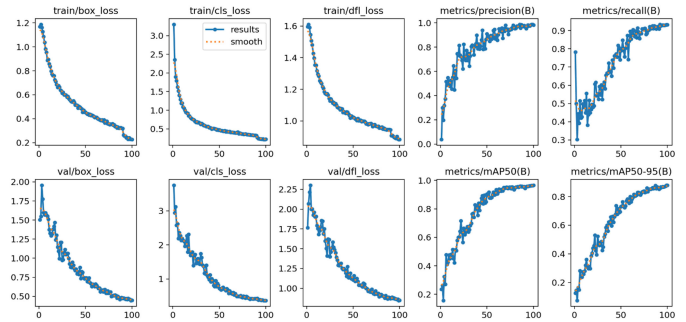
Trong quá trình huấn luyện, mô hình YOLOv11 đạt được các chỉ số lỗi như sau:

#### Tập huấn luyện:

- Box Loss: giảm từ 1.2 xuống 0.22
- Classification Loss: giảm từ 3.2 xuống 0.25
- DFL Loss: giảm từ 1.6 xuống 0.85

#### Tập kiểm định:

- Box Loss: giảm từ 1.9 xuống 0.45
- Classification Loss: giảm từ 3.5 xuống 0.50
- DFL Loss: giảm từ 2.25 xuống 0.85



Hình 4. Biểu đồ huấn luyện mô hình

### B. Chỉ số đánh giá hiệu suất

Kết quả đánh giá hiệu suất của mô hình như sau:

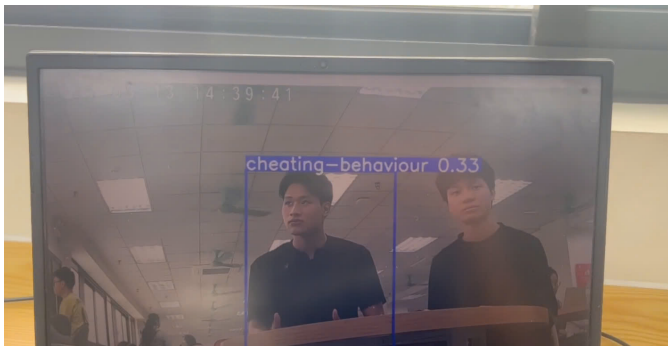
- Precision: 0.98
- Recall: 0.92
- mAP@0.5: 0.97
- mAP@0.5-0.95: 0.89

## X. KẾT QUẢ ĐẠT ĐƯỢC KHI CHẠY MÔ HÌNH

Sau khi huấn luyện và chạy thử nghiệm trên tập kiểm định, mô hình YOLOv11 đạt được những kết quả nổi bật như sau:

- **Độ chính xác:** YOLOv11 đạt độ chính xác cao trong việc phát hiện các hành vi gian lận, đặc biệt là lớp "cheating-behaviour" với độ chính xác (Precision) đạt 0.98 và Recall đạt 0.92.
- **Hiệu suất trên tập dữ liệu kiểm định:** Chỉ số mAP@0.5 đạt 0.97 và mAP@0.5-0.95 đạt 0.89, cho thấy mô hình có khả năng phát hiện hiệu quả và chính xác đối với nhiều trường hợp khác nhau.
- **Tốc độ xử lý:** YOLOv11 hoạt động với tốc độ nhanh, đáp ứng yêu cầu giám sát thời gian thực trong phòng thi. Thời gian xử lý mỗi khung hình là 15ms trên GPU, đảm bảo hệ thống có thể phát hiện gian lận ngay lập tức khi có sự kiện xảy ra.

Bên cạnh đó, hình minh họa kết quả phát hiện hành vi gian lận được đưa ra ở hình dưới, cho thấy mô hình YOLOv11 có thể phát hiện chính xác vị trí và phân loại hành vi gian lận trong phòng thi:



Hình 5. Biểu đồ huấn luyện mô hình

Như vậy, kết quả cho thấy mô hình YOLOv11 có thể được sử dụng hiệu quả trong việc giám sát và phát hiện các hành vi gian lận trong phòng thi, đáp ứng được yêu cầu về độ chính xác và tốc độ xử lý trong hệ thống thời gian thực.

## XI. ƯU ĐIỂM

Mô hình YOLOv11 sở hữu nhiều ưu điểm nổi bật, đặc biệt trong việc phát hiện hành vi gian lận trong phòng thi:

- **Độ chính xác cao:** Mô hình hoạt động xuất sắc với độ chính xác (Precision) đạt 0.98, đặc biệt trong việc phát hiện lớp "cheating-behaviour". Với Recall đạt 0.92, YOLOv11 có khả năng phát hiện chính xác các hành vi gian lận trong hầu hết các tình huống.
- **Hiệu suất huấn luyện và kiểm định:** Mô hình liên tục cải thiện trong suốt quá trình huấn luyện, với các chỉ số lỗi (Box Loss, Classification Loss, DFL Loss) đều giảm đáng kể, cho thấy quá trình tối ưu hiệu quả. Hiệu suất tổng thể của mô hình được đánh giá cao nhờ vào chỉ số mAP@0.5 đạt 0.97 và mAP@0.5-0.95 đạt 0.89 trên tập kiểm định.
- **Không có dấu hiệu overfitting:** Mô hình không có dấu hiệu overfitting rõ ràng trong quá trình huấn luyện. Sự ổn

định giữa tập huấn luyện và kiểm định giúp đảm bảo rằng mô hình có thể tổng quát hóa tốt khi áp dụng vào các tình huống thực tế.

- **Tốc độ xử lý nhanh:** YOLOv11 được tối ưu hóa để hoạt động với tốc độ xử lý nhanh, phù hợp với yêu cầu giám sát thời gian thực trong phòng thi. Mô hình có thể xử lý khoảng 67 FPS (khung hình trên giây) trên GPU, cho phép giám sát liên tục và đưa ra cảnh báo ngay lập tức khi phát hiện hành vi gian lận.
- **Khả năng ứng dụng rộng rãi:** Mô hình không chỉ dừng lại ở việc phát hiện hành vi gian lận trong thi cử mà còn có thể được triển khai cho nhiều ứng dụng khác nhau như giám sát an ninh, quản lý giao thông, và các lĩnh vực khác yêu cầu phân tích hình ảnh nhanh và chính xác.

## XII. HẠN CHẾ

Mặc dù YOLOv11 có nhiều ưu điểm, mô hình vẫn gặp phải một số hạn chế cần được giải quyết trong các ứng dụng thực tế:

- **Yêu cầu tài nguyên tính toán lớn:** Việc huấn luyện mô hình đòi hỏi lượng lớn tài nguyên tính toán, đặc biệt là yêu cầu cao về GPU. Trong môi trường không có GPU mạnh mẽ, quá trình huấn luyện có thể bị chậm lại, ảnh hưởng đến khả năng tối ưu và thời gian triển khai của mô hình.
- **Hiệu suất phát hiện trên một số lớp chưa cao:** Mặc dù hiệu suất trên lớp "cheating-behaviour" rất tốt, nhưng các lớp như "Normal" và "background" vẫn chưa đạt được độ chính xác tối ưu. Điều này có thể dẫn đến việc mô hình gặp khó khăn khi phân biệt giữa các hành vi gian lận và các hành vi bình thường trong một số trường hợp đặc thù.
- **Yêu cầu tối ưu thêm cho các tình huống đa dạng:** Để mô hình có thể tổng quát hóa tốt hơn, cần cải thiện thêm khả năng nhận diện trên các bối cảnh khác nhau. Việc phát hiện hành vi gian lận trong các phòng thi với ánh sáng, góc quay và chất lượng video khác nhau vẫn có thể là một thách thức.
- **Khó khăn trong triển khai hệ thống thời gian thực:** Mặc dù YOLOv11 có tốc độ xử lý nhanh, việc triển khai hệ thống giám sát thời gian thực ở quy mô lớn vẫn có thể gặp khó khăn nếu hệ thống không đủ mạnh về phần cứng và kết nối mạng. Hơn nữa, việc duy trì khả năng xử lý liên tục đòi hỏi sự đồng bộ tốt giữa phần mềm và phần cứng.

## TÀI LIỆU