



## Nghiên cứu phát triển hệ thống giám sát an ninh theo thời gian thực sử dụng thuật toán học sâu tiên tiến

## Research development of a real-time surveillance system using advanced deep learning algorithm

Hoàng Đức Quý<sup>1,\*</sup>, Cao Hữu Vinh<sup>2</sup>, Đoàn Văn Đồng<sup>3</sup>

<sup>1</sup>Viện CNTT và Điện, Điện tử, Nhóm nghiên cứu BRIDGE, Trường Đại học Giao thông vận tải TP. HCM

<sup>2</sup>Phòng Đào tạo, Trường Đại học Giao thông vận tải TP. HCM

<sup>3</sup>Viện Cơ khí, Nhóm nghiên cứu BRIDGE, Trường Đại học Giao thông vận tải TP. HCM

Từ khóa:

TÓM TẮT

Trí tuệ nhân tạo  
Hệ thống giám sát  
Kiến trúc YOLO  
Thuật toán ByteTrack

Bài báo này nghiên cứu và triển khai một hệ thống giám sát an ninh thông minh ứng dụng trí tuệ nhân tạo (AI) trong các môi trường giám sát khác nhau. Hệ thống tích hợp mô hình học sâu tiên tiến như YOLOv11 để phát hiện, theo dõi đối tượng theo thời gian thực, kết hợp giao diện người dùng trực quan phát triển bằng PyQt5. Các chức năng chính bao gồm: định nghĩa vùng cảnh báo (ROI), đếm người, thống kê và hiển thị dữ liệu. Kết quả thực nghiệm cho thấy hệ thống đạt độ chính xác cao (Precision và Recall > 90%), tốc độ xử lý ổn định (> 20 FPS), khả năng vận hành mượt mà, dễ sử dụng. Hệ thống được đánh giá có tính ứng dụng thực tiễn cao và tiềm năng mở rộng cho nhiều môi trường giám sát khác.

Keywords:

ABSTRACT

Artificial Intelligent  
Surveillance system  
YOLO  
ByteTrack

This paper presents the research and development of an AI-powered intelligent security monitoring system for various surveillance environments. The system integrates advanced deep learning model, YOLOv11, to detect and track objects in real time, with a user-friendly interface built using PyQt5. Key features include ROI definition, people counting, statistical reporting, and real-time display. Experimental results demonstrate high accuracy (Precision and Recall > 90%), stable processing speed (> 20 FPS), and intuitive usability. The system is considered highly feasible for real-world deployment and shows strong potential for expansion into other surveillance environments.

### 1. Giới thiệu

Trong thời đại công nghệ số phát triển mạnh mẽ, nhu cầu tự động hóa và tối ưu hóa trong công tác quản lý an ninh ngày càng trở nên cấp thiết. Trong bối

cảnh cách mạng công nghiệp 4.0, việc tích hợp trí tuệ nhân tạo (AI) vào các hệ thống giám sát an ninh đóng vai trò thiết yếu trong việc nâng cao hiệu suất hệ thống và đã trở thành một trong những lĩnh vực

\* Hoàng Đức Quý. Viện CNTT và Điện, Điện tử, Nhóm nghiên cứu BRIDGE, Trường Đại học Giao thông vận tải TP. HCM  
Email: [quyhd@ut.edu.vn](mailto:quyhd@ut.edu.vn)

<https://www.doi.org/10.55228/JTST140611>

Ngày nhận bài: 7/08/2025; Ngày nộp bài sửa: 25/8/2025 ; Ngày chấp nhận đăng: 15/11/2025

Ngày xuất bản trực tuyến: 15/11/2025

pISSN: 1859-4263; eISSN: 3030-4261

nghiên cứu được quan tâm hàng đầu trên toàn cầu [1], [2], [3], [4]. Theo báo cáo thị trường của Grand View Research, quy mô thị trường thị giác máy tính trong lĩnh vực an ninh và giám sát đạt 1.602,9 triệu USD năm 2024 và dự kiến tăng trưởng với hàng năm 21,1% đến năm 2030 [5].

Với mật độ khách hàng ra vào lớn và môi trường hoạt động liên tục, việc giám sát bằng phương pháp thủ công truyền thống như quan sát trực tiếp qua camera giám sát thường không hiệu quả, dễ bỏ sót và phụ thuộc nhiều vào yếu tố con người. Trước thực trạng đó, việc ứng dụng các công nghệ mới như trí tuệ nhân tạo (AI), học sâu (Deep Learning), và thị giác máy tính (Computer Vision) vào hệ thống giám sát đã mở ra hướng đi mới, hiện đại và hiệu quả hơn trong việc quản lý, phát hiện sớm các hành vi bất thường và nâng cao độ an toàn cho tài sản cũng như con người.

Điển hình như trong nghiên cứu của Fatma và cộng sự [2], nhóm tác giả đề xuất tích hợp mô hình trí tuệ nhân tạo tiên tiến YOLO trong hệ thống giám sát giao thông. Hệ thống đề xuất có thể đánh giá các video giám sát giao thông thời gian thực và nhận diện mật độ giao thông. Kết quả thử nghiệm cho thấy hệ thống đạt độ chính xác trung bình 92.4% trong việc nhận diện phương tiện giao thông. Nghiên cứu [6] đề xuất sử dụng YOLOv11 để nhận diện và theo dõi phương tiện giao thông và phát hiện tai nạn. Ở phương diện khác, Laura và cộng sự [7] đề xuất hệ thống giám sát thông minh tại cửa hàng bán lẻ. Hệ thống đề xuất sử dụng YOLOv11 để phát hiện đối tượng và bộ phân loại XGBoost để phân loại hành vi đối tượng là bình thường hay bất thường. Kết quả đánh giá cho thấy độ chính xác trung bình đạt 94.35%.

Ở một phương diện khác, Minh Long Hoang [8] đề xuất việc tích hợp AI vào phương tiện bay không người lái trong việc giám sát các sự kiện bất thường như cháy nổ, cướp bóc, xâm nhập bất hợp pháp. Tác giả đề xuất mô hình YOLOv8 kết hợp với bộ phân loại phân tầng để nhận diện đối tượng và phân loại các hành vi bất thường. Cùng hướng, Gang Chen và cộng sự đề xuất SGST-YOLOv8 [9], kết hợp với thuật toán theo dõi để giám sát người đi đường và phương tiện trong môi trường đại học, thuật toán đề xuất vượt trội hơn về độ chính xác so với mô hình chuẩn và có gánh nặng tính toán ít hơn.

Trong nước, nhóm nghiên cứu của Đặng Hoàng Phúc và cộng sự [10] đề xuất sử dụng mô hình YOLO

trong hệ thống giám sát xâm nhập tại môi trường công trường xây dựng, đường ray xe lửa. Kết quả nghiên cứu cho thấy độ chính xác của thuật toán YOLOv5n đề xuất đạt 91% với tốc độ xử lý đạt 12 khung hình trên giây. Mặc dù có kết quả đánh giá khả quan nhưng tập dữ liệu dùng trong nghiên cứu còn hạn chế. Một nghiên cứu khác trong [11] sử dụng mô hình YOLOv8, một phiên bản tiên tiến của họ mô hình phát hiện đối tượng YOLO, nhằm xây dựng hệ thống phát hiện hành vi leo rào trong thời gian thực. Dữ liệu huấn luyện gồm 1.000 hình ảnh và 20 đoạn video ngắn, được thu thập từ Internet và hiện trường mô phỏng tại ký túc xá sinh viên. Kết quả thực nghiệm cho thấy hệ thống đạt độ chính xác trung bình mAP đạt 0.87 trên tập train, 0.82 trên tập valid và 0.79 trên tập test, trong khi chỉ số Precision và Recall cũng dao động từ 0.72 đến 0.87, thể hiện hiệu năng ổn định của mô hình trong điều kiện thực tế. Tuy nhiên, nhóm tác giả cũng thừa nhận hạn chế về kích thước dữ liệu và ảnh hưởng của điều kiện môi trường.

Lấy cảm hứng từ các nghiên cứu trên, trong bài báo này, nhóm tác giả nghiên cứu việc xây dựng một giải pháp tích hợp AI vào quá trình giám sát. Cụ thể, hệ thống khảo sát hiệu năng của mô hình trí tuệ nhân tạo tiên tiến trong việc nhận dạng đối tượng tên là YOLOv11. Thuật toán đề xuất sẽ được tích hợp trong hệ thống giám sát an ninh thời gian thực. Điểm nổi bật của hệ thống là khả năng xử lý theo thời gian thực, giao diện người dùng thân thiện, cho phép các chức năng như quản lý camera trực tiếp, xác định vùng cảnh báo (ROI), ghi nhận dữ liệu, thống kê số lượng người ra vào. Ngoài ra, hệ thống cũng hỗ trợ nhiều tính năng hữu ích khác như chụp ảnh, quay video, tìm kiếm theo thời gian và hiển thị dữ liệu dạng biểu đồ, giúp việc phân tích và đánh giá trở nên trực quan và hiệu quả hơn.

Phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày phương pháp nghiên cứu bao gồm tổng quan về hệ thống giám sát, mô hình học sâu đề xuất và kết quả đánh giá định lượng, phần 3 trình bày kết quả khảo sát thực tế hệ thống đề xuất, làm rõ điểm mạnh, những rào cản trước khi đưa ra kết luận và hướng phát triển ở phần 4.

## 2. Xây dựng hệ thống giám sát an ninh tích hợp thuật toán AI tiên tiến

Hình 1 trình bày sơ đồ tổng quan hệ thống giám sát an ninh đề xuất. Đầu vào hệ thống là dữ liệu hình

ảnh thu nhận từ camera giám sát (CCTV), vị trí lắp đặt camera có thể tại cửa ra vào chính, khu vực hành lang và các vị trí trọng yếu trong môi trường giám sát nhằm bao phủ toàn diện khu vực giám sát. Số lượng camera có thể tùy chỉnh dựa vào phạm vi môi trường giám sát. Dữ liệu thu được có thể dưới dạng luồng video trực tiếp qua giao thức RTSP hoặc video lưu trữ cục bộ. Dữ liệu này được gửi tới máy chủ trung tâm (CCTV Server) để thực hiện tiền xử lý và chuyển tiếp tới mô hình nhận diện đối tượng dựa trên kiến trúc YOLOv11 - một mô hình phát hiện đối tượng hiện đại, được sử dụng để nhận diện người trong khung hình với độ chính xác và tốc độ xử lý thời gian thực. Các đối tượng được phát hiện sẽ được khoanh vùng bằng khung bao (bounding box) và xác định tọa độ. Tiếp đó, hệ thống thực hiện theo dõi sự di chuyển của các đối tượng này qua các vùng quan tâm (Region of Interest - ROI) được định nghĩa trước, như lối vào, lối ra hoặc các khu vực giới hạn. Việc theo dõi qua ROI cho phép hệ thống nhận biết chính xác hướng di chuyển, đồng thời thực hiện đếm số lượng người ra/vào từng khu vực cũng như phát hiện các hành vi bất thường phục vụ mục tiêu giám sát và thống kê.

### 2.1. Thuật toán YOLOv11

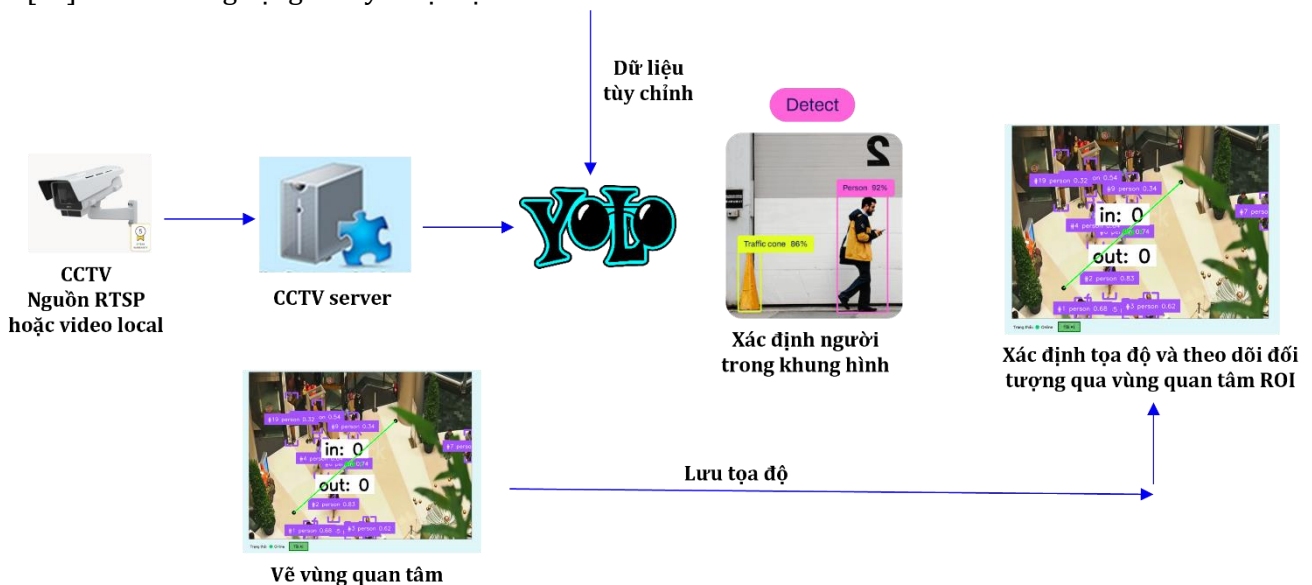
Trong thị giác máy tính hiện đại, YOLO là kiến trúc đột phá liên tục cải tiến trong thập kỷ qua. Phiên bản mới nhất, YOLOv11, vượt trội về cấu trúc và hiệu năng so với các bản trước như YOLOv5, YOLOv8 [12], [13]. Mô hình ứng dụng các kỹ thuật học sâu để tối ưu

hoá phát hiện, phân loại, phân đoạn và theo dõi đối tượng theo thời gian thực, với độ chính xác và tốc độ ấn tượng. Kết quả trên tập COCO cho thấy YOLOv11 vẫn đạt độ chính xác cao ngay cả với mô hình nhỏ. Tính linh hoạt và dễ triển khai giúp nó trở thành giải pháp hiệu quả cho các ứng dụng thị giác máy tính thực tế. Kiến trúc của YOLOv1 được minh họa trên Hình 3.

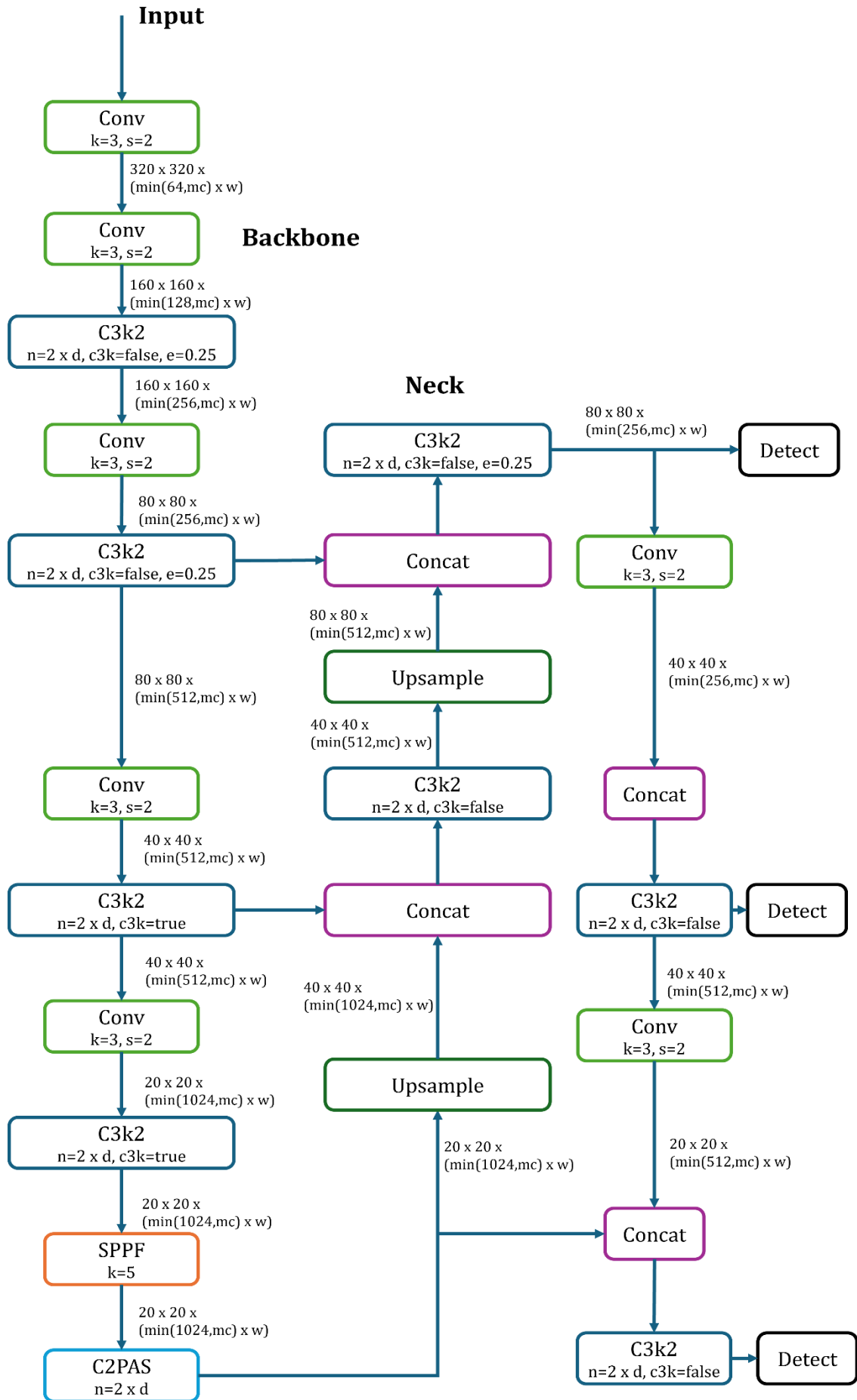
Kiến trúc tổng thể YOLOv11 gồm ba phần chính:

- Backbone: Trích xuất đặc trưng từ ảnh đầu vào qua chuỗi các khối học sâu;
- Neck: Tổng hợp đặc trưng đa tỷ lệ, tăng cường thông tin không gian cho việc phát hiện đối tượng ở nhiều kích cỡ khác nhau;
- Head: Thực hiện dự đoán các hộp giới hạn cho từng đối tượng trong ảnh. Mỗi hộp giới hạn chứa tọa độ vùng chứa đối tượng trên khung hình.

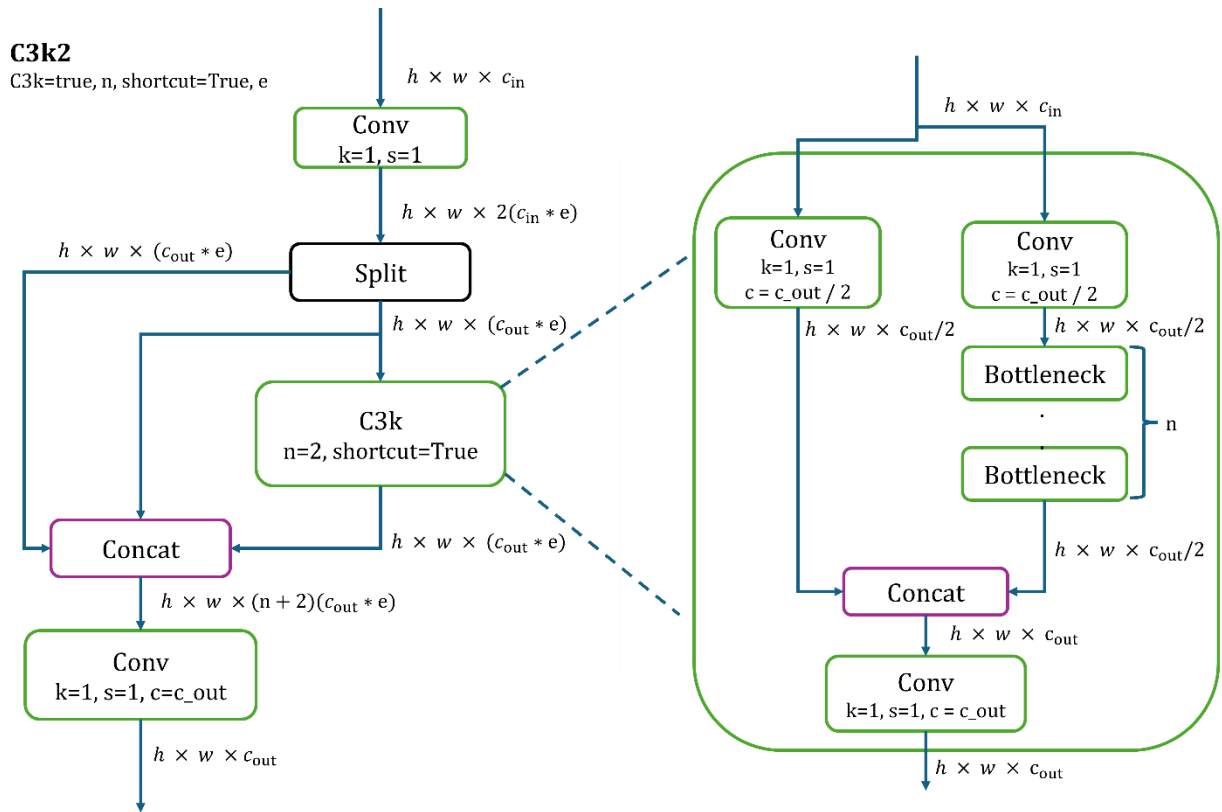
Cả ba phần đều có các khối đặc biệt được tối ưu hóa về tính toán và khả năng biểu diễn đặc trưng so với các phiên bản YOLO trước. Thành phần cơ bản của YOLOv11 là khối C3k2 kết hợp nhiều khối tích chập và khối bottleneck cho phép việc học các đặc trưng phức tạp (Hình 3). Khối C3k2 gồm 3 khối tích chập và một vài khối cổ chai được quyết định bởi tham số  $n$ . Theo nhóm tác giả, khối C3k2 có hiệu quả trong việc nắm bắt các đặc trưng phức tạp và đặc biệt thích hợp trong việc phát hiện các đối tượng đa kích thước.



Hình 1. Sơ đồ tổng quan hệ thống giám sát tích hợp AI đề xuất.



Hình 2. Kiến trúc mô hình YOLOv11.



Hình 3. Khối C3k2.

## 2.1. Thiết lập thí nghiệm

### 2.2.1. Tập dữ liệu và tiền xử lý

Tập dữ liệu đào tạo được thu thập trên Kaggle [14] với tất cả 17300 ảnh chuyên dùng cho tác vụ nhận dạng con người. Trong quá trình tiền xử lý dữ liệu, chúng tôi loại bỏ 300 ảnh không phù hợp (độ phân giải thấp, vỡ hình, không chứa đối tượng con người). Do đó, tập dữ liệu sau xử lý có 13000 ảnh (~76%) được sử dụng để đào tạo và 4000 (~24%) ảnh được dùng để đánh giá mô hình huấn luyện. Về việc lựa chọn tỉ lệ phân bổ, ước lượng này dựa trên thực tiễn phổ biến trong lĩnh vực thị giác máy tính, nhằm cân bằng giữa đủ dữ liệu huấn luyện cho mô hình và đảm bảo độ tin cậy của bộ đánh giá. Các ảnh trong tập dữ liệu này có sự đa dạng đáng kể về điều kiện môi trường, bao gồm:

Đa dạng về ánh sáng: Hình ảnh được chụp dưới các điều kiện ánh sáng khác nhau, từ ánh sáng ban ngày rõ ràng đến ánh sáng yếu hay có bóng đổ, giúp mô hình học được khả năng nhận diện tốt trong nhiều điều kiện ánh sáng thực tế.

Đa dạng về góc nhìn và phối cảnh: Hình ảnh được ghi lại từ nhiều góc quay, phối cảnh khác nhau, bao

gồm cả góc trực diện và góc nghiêng, nhằm nâng cao khả năng nhận dạng đối tượng từ nhiều hướng khác nhau.

Đa dạng về đối tượng và phong nền: Hình ảnh bao gồm các đối tượng với đặc điểm khác nhau (cơ người, tư thế, trạng thái) cùng nhiều phong nền phong phú, giúp cải thiện khả năng tổng quát hóa của mô hình khi áp dụng trong các môi trường giám sát khác nhau.

### 2.2.2. Thông số thiết lập

Nghiên cứu được thực hiện trên một máy chủ trang bị GPU NVIDIA 3090, 24 GB RAM, hệ điều hành ubuntu 24.4, Pytorch 2.5.1. Ảnh đầu vào được thay đổi kích thước đồng nhất 640 x 640. Kích thước lô ảnh được đặt là 16. Thuật toán huấn luyện là Adam với tốc độ học khởi tạo 3e - 4. Các mô hình được huấn luyện với tổng cộng 16240 bước lặp.

Các tiêu chí đánh giá bao gồm mAP@50, Precision và Recall Công thức 1 và Công thức 2. Trong đó, TP, FP, FN lần lượt là dương tính thật, dương tính giả và âm tính giả. Precision là tỉ lệ giữa số lượng các trường hợp dự đoán đúng là dương tính và tổng số lượng các trường hợp được dự đoán là dương tính. Recall là tỉ lệ giữa số lượng các trường hợp dự

đoán đúng là dương tính và tổng số lượng các trường hợp thực sự là dương tính.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

Ba mô hình thí nghiệm bao gồm YOLOv11-tiny, YOLOv11-small và YOLOv11-large.

### 2.2.3. Kết quả huấn luyện mô hình

Trong ba mô hình thí nghiệm, mô hình YOLOv11-small cho kết quả tổng thể tốt nhất (Xem Bảng 1) với mAP@50 đạt 84.3 (%), Precision đạt 89.2 (%) và Recall đạt 81.1 (%). Bên cạnh hiệu suất phát hiện đối tượng, YOLOv11-small còn đạt tốc độ xử lý 18 ms/ảnh, phù hợp cho các tác vụ yêu cầu xử lý thời gian thực như phân tích video. So sánh với nghiên cứu trước đó của nhóm tác giả trong [10] và [15], YOLOv11 có chỉ số chính xác trung bình thấp hơn nhưng có tốc độ xử lý vượt trội. Ngoài ra, số lượng mẫu đánh giá của nghiên cứu này cũng cao hơn 10 lần so với nghiên cứu trước đó nên có thể kết luận mô hình đề xuất có độ khái quát cao hơn. Từ các lý do trên, chúng tôi chọn YOLOv11-small để tích hợp vào hệ thống giám sát đề xuất.

### 2.2. Thuật toán theo dõi và đếm số lượng vào ra

Ở bài báo này, sau khi mô hình học sâu nhận diện được đối tượng. Thuật toán ByteTrack [16] được sử dụng để theo dõi người vào/ ra trong khung hình. Cụ thể, với mỗi khung hình  $I_t$ , thuật toán đề xuất có các bước thực thi như sau:

**Thuật toán 1** Thuật toán theo dõi và đếm số lượng vào ra.

1	Xác định số người trong khung:	(3)
	$D_t = YOLO(I_t)$	
2	Liên kết các phát hiện với các thẻ theo dõi:	(4)
	$T_t = ByteTrack(D_t)$	
3	Định nghĩa đường đếm (Region of Interest - ROI):	(5)
	$L: ax + by + c = 0$	
4	Xác định đối tượng bằng qua ROI và hướng di chuyển:	

$$L(p_{t-1}) \cdot L(p_t) < 0 \quad (6)$$

$$\begin{cases} C_{in} = C_{in} + 1, L(p_{t-1}) > 0, L(p_t) < 0 \\ C_{out} = C_{out} + 1, \text{ngược lại} \end{cases} \quad (7)$$

Trong đó,  $D_t = \{d_{t,1}, d_{t,2}, \dots, d_{t,n}\}$  là tập hợp các đối tượng phát hiện được bởi mô hình YOLO tại thời điểm  $t$ .  $T_t = \{T_{t,1}, T_{t,2}, \dots, T_{t,k}\}$  là tập hợp các thẻ theo dõi gắn liền với đối tượng trong khung hình tại thời điểm  $t$ . Mỗi thẻ theo dõi chứa một ID riêng và danh sách tọa độ của đối tượng từng khung ảnh.  $C_{in}, C_{out}$  lần lượt là biến đếm khi có đối tượng vào ra.  $p_{t-1}, p_t$  là tọa độ tâm của khung bao đối tượng trong hai khung ảnh liên tiếp.

### 3. Kết quả đánh giá thực nghiệm và bàn luận

Để đánh giá hiệu quả của hệ thống giám sát đề xuất, nhóm tác giả đã xây dựng một quy trình thực nghiệm với dữ liệu ghi hình từ hệ thống camera giám sát. Cụ thể, bộ dữ liệu bao gồm 5 đoạn video thực tế, được trích xuất từ camera CCTV gắn tại vị trí khu vực cửa ra vào. Mỗi video có thời lượng từ 2 đến 3 phút, tổng cộng đạt 12 phút 30 giây, với độ phân giải HD. Các video có bối cảnh ghi hình khác nhau từ trung tâm thương mại, hành lang công ty với các điều kiện môi trường khác biệt về cường độ ánh sáng, mật độ di chuyển, tình huống gây nhiễu như ma-nơ-canh trong trung tâm thương mại, bóng đổ, người bị che khuất.

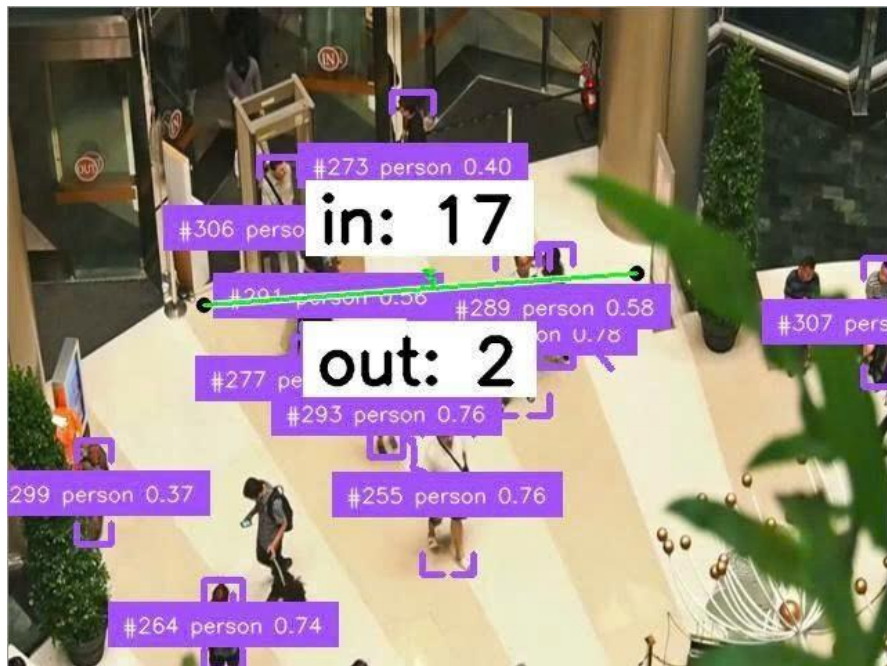
**Bảng 1.** So sánh hiệu suất các mô hình đề xuất và các công trình gần đây trong tác vụ phát hiện đối tượng.

Tên mô hình	Tiêu chí đánh giá			
	mAP@50 (%)	Precision (%)	Recall (%)	Speed (ms/ảnh)
YOLOv2 [15]	96.8	-	-	~510
YOLOv8-n [10]	92.2	-	-	~66
YOLOv8-s [10]	92.3	-	-	~125
YOLOv8-m [10]	92.1	-	-	~333
YOLOv11-n	78.9	86.8	77.6	~13
YOLOv11-s	84.3	89.2	81.1	~18
YOLOv11-m	85.7	90.4	88.6	~37

**Bảng 2.** Kết quả đánh giá thực nghiệm hệ thống đề xuất (Các thí nghiệm được thực hiện trên một máy tính xách tay có CPU Intel core I5-1135G7, RAM 16GB, không GPU).

Video	Tiêu chí đánh giá					
	Thực tế (số người)	Dự đoán (số người)	Sai số (%)	Precision (%)	Recall (%)	Tốc độ (frame/s)
01	22	21	4.5	95	91	29
02	18	17	5.5	93	89	27
03	28	25	7.4	92	88	25
04	15	15	0.0	97	94	32
05	12	13	8.3	94	90	30

a)



b)



**Hình 4.** Các trường hợp sai sót của hệ thống. (a) Thể hiện trường hợp đếm thiếu đối tượng khi mật độ đông đúc; (b) Thể hiện trường hợp đếm nhầm vật thể là đối tượng.

Bảng 2 tổng hợp kết quả đánh giá hệ thống trên các tiêu chí Độ chính xác, Độ nhạy, Sai số, và Tốc độ xử lý trên CPU. Kết quả đánh giá định lượng cho thấy hệ thống đề xuất có sai số dao động từ 0 ~ 8.3%. Độ chính xác trung bình đạt 94.2% và độ nhạy trung bình là 90.4%. Tốc độ xử lý dao động từ 25 đến 32 (khung hình/giây). Các sai sót của hệ thống chủ yếu đến từ các trường hợp khó như khi số lượng đối tượng quá đông băng qua vùng giám sát, hay nhầm lẫn vật thể là đối tượng và nhầm lẫn bóng là đối tượng (Hình. 4). Từ các kết quả đánh giá, nhóm nghiên cứu nhận định hiệu suất của hệ thống là tương đối cao, có tiềm năng tích hợp vào hệ thống thực tế với một vài tinh chỉnh phù hợp. Tuy nhiên, độ phức tạp của môi trường giám sát, chất lượng video và phần cứng tính toán sẽ có ảnh hưởng đáng kể đến hiệu suất của hệ thống. Do đó, một số cải tiến như thêm ngữ cảnh huấn luyện, tối ưu thông số mô hình là một số phương hướng cần được tập trung trong tương lai.

#### 4. Kết luận

Trong bài báo này, nhóm nghiên cứu đề xuất hệ thống giám sát an ninh tích hợp thuật toán AI tiên tiến YOLOv11 trong việc nhận dạng và theo dõi các đối tượng con người. Các kết quả đánh giá định lượng cho thấy độ chính xác và độ nhạy của hệ thống đạt hiệu suất cao, trên 90%. Bên cạnh đó, tốc độ suy luận của hệ thống cũng đạt xấp xỉ trung bình 28 FPS trên một máy tính cục bộ cho thấy khả năng áp dụng vào thực tế của mô hình. Tuy nhiên, kết quả đánh giá cũng chỉ ra những hạn chế về ảnh hưởng nhiễu từ môi trường giám sát, chất lượng video, phần cứng tính toán.

Trong tương lai, nhóm nghiên cứu sẽ tập trung vào việc bổ sung dữ liệu huấn luyện, tinh chỉnh thông số mô hình và thêm vào các kỹ thuật nhận diện đối tượng tiên tiến để giải quyết các hạn chế trên. Bên cạnh đó, hệ thống giám sát đề xuất sẽ được tích hợp thêm các tính năng:

(i) Nhận dạng và theo dõi chuyển động liên tục của các đối tượng trong vùng giám sát, giúp phát hiện các hành vi bất thường như dừng lại lâu, xâm nhập khu vực cấm hoặc di chuyển ngược chiều;

(ii) Cảnh báo tự động khi phát hiện các hành vi nghi vấn dựa trên dữ liệu chuyển động và phân tích mô hình học sâu, tăng cường an ninh chủ động.

#### Đóng góp của các tác giả trong bài báo

**Hoàng Đức Quý:** Phương pháp, Quản lý dữ liệu, Phân tích chính thức, Trực quan hóa, Phản hồi ý kiến phản biện, Viết – bản thảo gốc. **Cao Hữu Vinh:** Phân tích dữ liệu, Viết – bản thảo gốc. **Đoàn Văn Đồng:** Giám sát, chỉnh sửa bản thảo.

#### Tuyên bố không xung đột lợi ích và cam kết bản quyền

Chúng tôi cam kết nghiên cứu này không có xung đột lợi ích và các kết quả nghiên cứu là độc nhất chưa từng công bố dưới mọi hình thức.

#### Chia sẻ dữ liệu theo yêu cầu

Dữ liệu không được cung cấp theo yêu cầu

#### Lời cảm ơn

Chúng tôi cảm ơn nhóm nghiên cứu BRIDGE đã hỗ trợ cơ sở vật chất cho nghiên cứu.

1<sup>st</sup> Hoang Duc Quy\*. *Big Data Research for Infrastructure & Green Engineering (BRIDGE), Ho Chi Minh City University of Transport*

2<sup>nd</sup> Cao Huu Vinh. *Training Department, Ho Chi Minh City University of Transport*

3<sup>rd</sup> Doan Van Dong. *Big Data Research for Infrastructure & Green Engineering (BRIDGE), Ho Chi Minh City University of Transport*

\*Corresponding author: quyhd@ut.edu.vn

#### Tài liệu tham khảo

- [1] P. R. Kalluri, W. Agnew, M. Cheng, K. Owens, L. Soldaini, and A. Birhane, "Computer-vision research powers surveillance technology," *Nature*, vol. 643, no. 8070, pp. 73–79, Jul. 2025, Doi: 10.1038/s41586-025-08972-6.
- [2] F. M. Talaat, R. M. El-Balka, S. Sweidan, S. A. Gamel, and A. M. Al-Zoghby, "Smart traffic management system using YOLOv11 for real-time vehicle detection and dynamic flow optimization in smart cities," *Neural Comput Appl*, 2025, Doi: 10.1007/s00521-025-11434-9.
- [3] M. Cabanillas-Carbonell, J. Sallari Rivera, and J. Santivañez Muñoz, "Artificial intelligence in video surveillance systems for suspicious activity detection and incident response: A systematic review," *Advances in Science and Technology. Research Journal*, vol. 19, no. 3, 2025.
- [4] C. Fontes, E. Hohma, C. C. Corrigan, and C. Lütge, "AI-powered public surveillance systems: why we (might) need them and how we want them," *Technol Soc*, vol. 71, p. 102137, 2022.

- [5] Grand Review Research Inc., "Security & Surveillance - Computer Vision Market Statistics," 2025. [Online]. Available: <https://www.grandviewresearch.com/horizon/statistics/computer-vision-market/non-industrial/security-surveillance/global>. [Accessed: Jul. 30, 2025].
- [6] R. R. F. Lupian, C. G. Arong, W. S. Betinol, and D. B. Valdez, "Intelligent Traffic Monitoring And Accident Detection System Using YOLOv11 And Image Processing," in *2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 2025, pp. 1–5, Doi: 10.1109/eStream66938.2025.11016862.
- [7] L. Tsanaullailla, F. Hamami, and S. Suakanto, "Smart Surveillance in Retail Stores: Detection, Classification, and Documentation of Suspicious Behavior Utilizing YOLOv11 and XGBoost," in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, 2025, pp. 127–132, Doi: 10.1109/ICERA66156.2025.11087345.
- [8] M. L. Hoang, "Smart Drone Surveillance System Based on AI and on IoT Communication in Case of Intrusion and Fire Accident," *Drones*, vol. 7, no. 12, Dec. 2023, Doi: 10.3390/drones7120694.
- [9] G. Cheng, P. Chao, J. Yang, and H. Ding, "SGST-YOLOv8: an improved lightweight YOLOv8 for real-time target detection for campus surveillance," *Applied Sciences*, vol. 14, no. 12, p. 5341, 2024.
- [10] Đ. H. Phúc, L. Anh, U. Vũ, and H. Đức Quý, "Research and development of an intrusion warning system using advanced artificial intelligence algorithms," *Tạp chí khoa học công nghệ Giao thông vận tải*, vol. 13, no. 1, 2023.
- [11] N. Đức Lưu, N. Văn Trọng, and L. Văn Út, "Phát hiện hành vi leo rào qua camera giám sát bằng giải thuật học sâu climbing behavior detection via surveillance camera by using deep learning," *Tạp chí khoa học Trường Đại học Bạc Liêu BLUNI Journal of Science*, vol. 6, no. 12, pp. 77–87, 2024.
- [12] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review A PREPRINT," 2025, Doi: 10.48550/arXiv.2501.13400.
- [13] M. Sohan, T. Sai Ram, and Ch. V. Rami Reddy, "A Review on YOLOv8 and Its Advancements," in *Data Intelligence and Cognitive Informatics (ICDICI2023)*, 2024, pp. 529–545, Doi: 10.1007/978-981-99-7962-2\_39.
- [14] Fares Elmenshawii, "Human Dataset," 2023. [Online]. Available: <https://www.kaggle.com/datasets/fareselmenshawii/human-dataset>. [Accessed: Jul. 31, 2025].
- [15] H. H. Nguyen, T. N. Ta, N. C. Nguyen, V. T. Bui, H. M. Pham, and D. M. Nguyen, "Yolo based real-time human detection for smart video surveillance at the edge," in *2020 IEEE eighth international conference on communications and electronics (ICCE)*, IEEE, 2021, pp. 439–444.
- [16] Y. Zhang *et al.*, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in *European Conference on Computer Vision (ECCV2022)*, Apr. 2022, Doi:10.1007/978-3-031-20047-2\_1.