

CẢI THIỆN HIỆU QUẢ PHÁT HIỆN HÀNH VI MANG VŨ KHÍ**IMPROVE DETECTION PERFORMANCE OF WEAPONS CARRYING BEHAVIOR****Đoàn Thị Hương Giang**

Trường Đại học Điện lực

Ngày nhận bài: 20/12/2023, Ngày chấp nhận đăng: 22/02/2024, Phản biện: TS. Nguyễn Thị Thanh Nhân

Tóm tắt:

Phát hiện hành vi mang vũ khí là một trong những bài toán cấp bách và yêu cầu triển khai trên phạm vi rộng với nhiều camera và thường yêu cầu cài đặt trên cấu hình hệ thống gọn nhẹ nhưng cần phải đạt được độ chính xác cao. Mạng YOLO (You Only Look Once) là một trong những mô hình phát hiện đối tượng trong ảnh khá hiệu quả hiện nay. Tuy nhiên, các mô hình YOLO được công bố cho cộng đồng dùng chung không được huấn luyện trên các cơ sở dữ liệu (CSDL) vũ khí. Nghiên cứu này thực hiện thu thập và gán nhãn bộ CSDL của 06 loại vũ khí thông dụng. Ngoài ra, bài báo cũng triển khai giải pháp học chuyển giao giữa các mô hình YOLO cấu hình cao và nặng cho các mô hình YOLO cấu hình thấp và nhẹ hơn. Việc này giúp chuyển giao tri thức mà hệ chuyên gia đã tổng hợp được đối với các loại vũ khí cho các mô hình nhẹ nhằm đạt được một mô hình có cấu hình đơn giản, dễ dàng triển khai thực tế trong khi độ chính xác hệ thống cao hơn. Nghiên cứu còn đề xuất giải pháp phát hiện hành vi mang vũ khí của người sử dụng kết hợp kết quả đầu ra của mô hình YOLO với luồng dịch chuyển quang học của các đối tượng phát hiện trong ảnh. Kết quả đạt được cho thấy hệ thống của chúng tôi đạt kết quả tốt hơn khi học chuyển giao để phát hiện vũ khí và đặc biệt đạt kết quả cao hơn khi phát hiện hành vi mang vũ khí. Độ chính xác thử nghiệm khi phát hiện vũ khí khi học chuyển giao giữa mô hình YOLO V3 tiny so với YOLO V3 tiny – YOLO V8 KD cao hơn từ 2% đến 7,9% trong khi thời gian đáp ứng nhanh hơn 10,95 ms khi chạy trên GPU. Độ chính xác phát hiện hành vi mang vũ khí của giải pháp đề xuất được cải thiện lên tới 13,88% so với sử dụng kết quả đầu ra của mô hình của YOLO V3 tiny.

Từ khóa:

Mạng nơ-ron tích chập, học sâu, chắt lọc kiến thức, mô hình giáo viên – sinh viên, học chuyển giao, phát hiện người, phát hiện vật thể.

Abstract:

Weapon carrying behavior detection is one of the necessary problems. This system deploys on a large space with many cameras. It is installed in a compact and slow configuration while it requires high accuracy. The YOLO network is one of the most effective object detection models in image/video. However, the published YOLO (You Only Look Once) models are not trained on weapons databases. This study collects and labels the database of 6 common types of weapons. In addition, the paper also implements a transfer learning solution between high-configuration and heavy YOLO models to lower-configuration and smaller YOLO models. This work aims to transfer the knowledge from the expert system that has compiled about weapons to a simpler model to achieve a model with a simple configuration, easy to deploy in practice while maintaining higher accuracy. The research also proposes a solution to detect the user's weapon carrying behavior that combines the

output of the YOLO model with the optical flow stream of detected humans and weapons in the image. The results show that our system achieves better results when transferring learning to detect weapons and especially achieves better results when detecting weapon carrying behavior. Experimental weapon detection accuracy between YOLO V3 tiny models compared to YOLO V3 tiny – YOLO V8 KD is 2% to 7.9% higher while response time is 10.95 ms faster when tested on GPU. The weapon carrying behavior detection accuracy of the proposed solution is improved by up to 13.88% compared to using the model output of YOLO V3 tiny.

Keywords:

Convolution Newral Network, Deep Learning, Weapon Detection, Transfer Learning, Human Detection, Object Detection.

1. GIỚI THIỆU

Phát hiện hành vi mang vũ khí là một trong những bài toán cấp thiết trong lĩnh vực an ninh giúp phát hiện sớm và ngăn ngừa hậu quả nghiêm trọng có thể xảy ra cho con người. Đây là một trong những nhiệm vụ cấp bách nhằm đảm bảo xã hội an toàn và nâng cao chất lượng cuộc sống. Tuy nhiên, với các bài toán an ninh có yếu tố xã hội thường triển khai trên hệ thống mạng camera tại nhiều điểm trên đường phố, khu vui chơi, khu thể thao, công viên,... Do đó, hệ thống được xử lý trên một số lượng lớn camera và đòi hỏi hệ thống phải có thời gian đáp ứng nhanh, cấu hình nhẹ và độ chính xác phát hiện trong chỉ giới an toàn cho phép. Trong khi đó, mạng nơron tích chập hiện nay thể hiện rõ ưu điểm vượt trội trong nhiều lĩnh vực khác nhau như xử lý ảnh [1], xử lý ngôn ngữ tự nhiên [2],... Đặc biệt, mạng YOLO [3]-[10] là một trong những mạng tích chập phát hiện tốt các đối tượng trong ảnh. Các phiên bản YOLO V1 [3] đầu tiên được đề xuất năm 2016 và cho đến

nay các phiên bản YOLO V8 [10] mới nhất được cập nhật năm 2023 nhằm cải tiến các phiên bản trước đó. Các phiên bản càng nâng cấp thì độ chính xác càng cải thiện nhưng yêu cầu cấu hình mạng càng phức tạp và do vậy cần cài đặt trên hệ thống máy tính có cấu hình cao hơn hoặc là thời gian đáp ứng lâu hơn. Điều đó không chỉ với các phiên bản khác nhau của YOLO mà còn đối với cả các mô hình mạng nơron tích chập khác như Resnet [11], DenseNet [12], MobileNet [13]. Điều đó là do các phiên bản sau có cấu hình phức tạp hơn nên số lượng tham số lớn, tri thức mà mạng tổng hợp được tốt hơn so với cấu hình đơn giản hơn. Do vậy, học chuyển giao và chắt lọc tri thức [14] (Knowledge Distillation) đã được các nhà khoa học nghiên cứu và đề xuất lần đầu từ những năm 2005 nhằm sử dụng mô hình chuyên gia có nhiều kiến thức để truyền cho mô hình đơn giản có tri thức kém hơn khi thực hiện huấn luyện và tinh chỉnh lại tham số của mô hình đơn giản trên cùng một bộ CSDL. Tuy nhiên, việc

học chuyển giao trên mô hình YOLO cho bài toán phát hiện vũ khí thì chưa có nghiên cứu nào thực hiện. Do đó, trong nghiên cứu này sẽ khảo sát quá trình học chuyển giao từ hệ chuyên gia là các mô hình YOLO có cấu hình cao như YOLO V6 [8], YOLO V7 [9] và YOLO V8 [10] cho mô hình có cấu hình thấp là YOLO V3 tiny [5] trên CSDL vũ khí.

Hiện nay, có một số bộ CSDL vũ khí được các nhóm nghiên cứu thu thập, gán nhãn và chia sẻ cho cộng đồng dùng chung như Pistol classification dataset [15], Knife classification dataset [16], Sohas weapon dataset [17], Pistol detection dataset [18], Knife detection dataset [19], Sohas weapon detection dataset [20]. Tuy nhiên, các CSDL này có một số hạn chế chính đó là: (1) Hầu hết chỉ có số lượng ít các loại vũ khí gồm súng, dao, côn; (2) Các bộ CSDL có điều kiện nền đơn giản, hầu hết chỉ chứa hình ảnh vũ khí; (3) Chúng không có hình ảnh và gán nhãn về hành vi mang vũ khí; (4) Các bộ CSDL có hình ảnh tĩnh; (5) CSDL chứa lẫn với cả vũ khí và các vật dụng thông thường khác như ví tiền,... Do đó, trong nghiên cứu này chúng tôi đã định nghĩa 06 loại vũ khí khác nhau. Sau đó, chúng tôi thực hiện thu thập các video người mang vũ khí và gán nhãn cho bộ CSDL này trong điều kiện nền phức tạp ở môi trường ngoài trời và môi trường trong nhà.

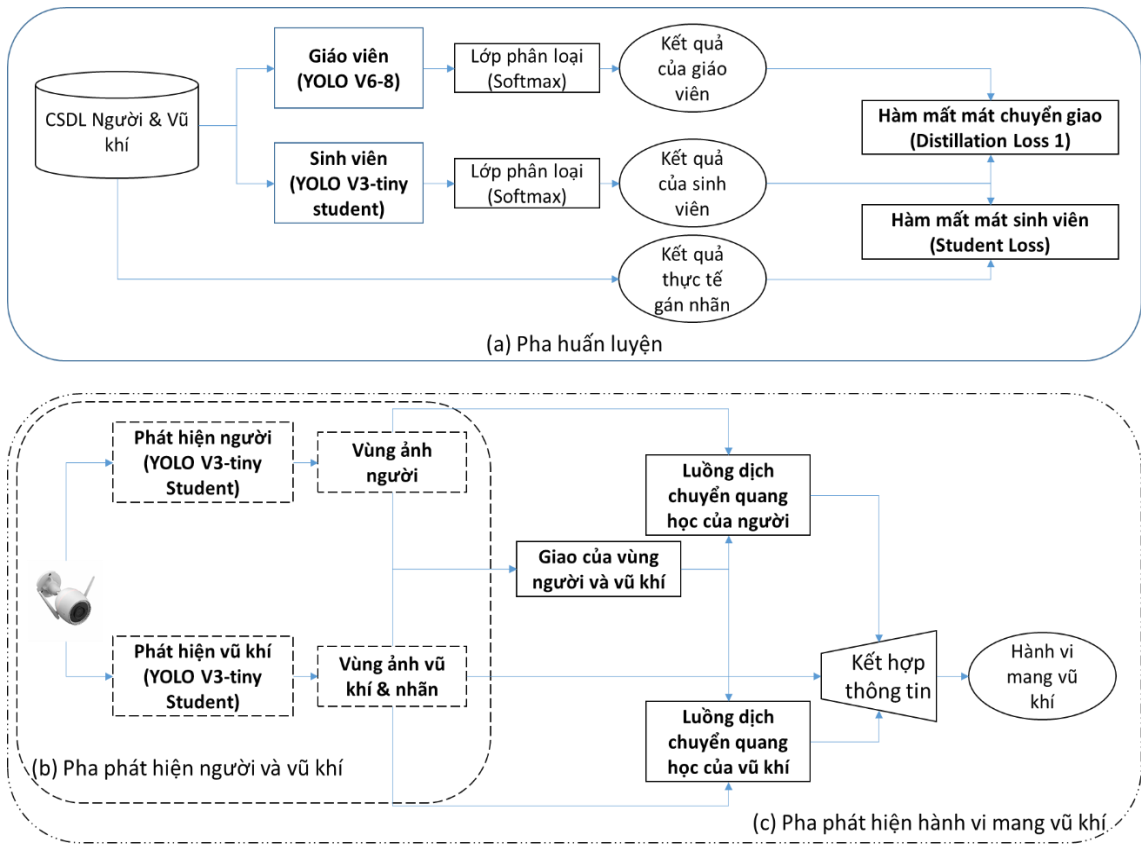
Ngoài ra, việc sử dụng mô hình YOLO chỉ có thể phát hiện vũ khí mà không thể phát hiện hành vi có mang/vác/đeo vũ khí của người. Trong nghiên cứu này chúng tôi đề xuất giải pháp phát hiện hành vi mang vũ khí của người kết hợp giữa kết quả đầu ra của mô hình YOLO với luồng dịch chuyển quang học [21] của đối tượng trong ảnh khi phát hiện được từ mạng YOLO. Giải pháp đề xuất được thực hiện thử nghiệm trên hai bộ CSDL gồm 01 bộ CSDL dùng chung Pistol detection dataset [18] và 01 bộ CSDL tự thu thập. Kết quả thử nghiệm cho thấy giải pháp đề xuất hiệu đạt hiệu quả cao hơn so với giải pháp gốc cả về độ chính xác và thời gian đáp ứng.

Phần 2 của bài báo 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

2.1. Phát hiện hành vi mang vũ khí của người

Hệ thống đề xuất gồm ba phần chính: (1) Quá trình huấn luyện hệ thống sử dụng mô hình chất lọc và chuyển giao tri thức (Hình 1a); (2) Quá trình phát hiện vùng người và vùng vũ khí (Hình 1b); (3) Quá trình phát hiện hành vi mang vũ khí của người (Hình 1c). Ba quá trình trên được trình bày chi tiết trong các phần tiếp theo.



Hình 1. Sơ đồ khối hệ thống phát hiện hành vi mang vũ khí của người
(a) Huấn luyện mô hình; (b) Hệ thống phát hiện người và vũ khí; (c) Phát hiện hành vi mang vũ khí.

2.2.1. Huấn luyện mô hình

Mạng YOLO được lựa chọn để huấn luyện mô hình phát hiện vùng người và 06 loại vũ khí. Bộ CSDL này được mô tả trong mục 2.2. Tuy nhiên, để có mô hình đạt hiệu suất cao trong khi kiến trúc mô hình nhẹ hơn thì chúng tôi sử dụng mô hình chất lọc tri thức khi học chuyển giao như mô tả trong Hình 1a. Trong đó, mô hình chất lọc tri thức chuyển giao giữa các mô hình YOLO trong Hình 1a được thực hiện thông qua hàm mất mát là sự kết hợp của hàm mất mát của sinh viên L_s (Student Loss) và hàm mất mát kết hợp

giữa hai giáo viên và sinh viên L_{KD} (Distillation Loss) như mô tả trong công thức (1) sau đây:

$$L(y, \hat{y}^s, \hat{y}^t) = L_s(y, \hat{y}^s) + \alpha L_{KD1}(\hat{y}^t, \hat{y}^s) \quad (1)$$

Trong đó, $0 \leq \alpha \leq 1$ hệ số thể hiện sự ảnh hưởng của hàm mất mát của sinh viên L_s và hàm mất mát kết hợp giữa hai giáo viên L_{KD} . Trong bài báo này, hệ số α sử dụng bằng 0.5. Ngoài ra, hàm mất mát sử dụng là CE (Cross Entropy) [22], và công thức (1) được biểu diễn chi tiết như trong công thức (2) sau đây:

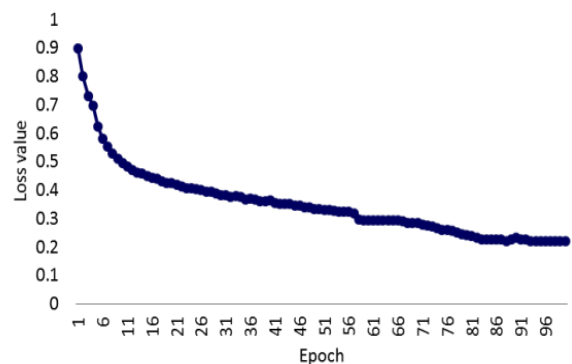
$$L(y, \widehat{y^s}, \widehat{y^t}) = \sum_i^N (y_i \log \widehat{y_i^s}) + \alpha \sum_i^N (\widehat{y_i^t} \log \widehat{y_i^s}) \quad (2)$$

Triển khai chi tiết hệ thống được mô tả trong Bảng 1. Trong đó, mô hình giáo viên sử dụng lần lượt một trong ba phiên bản YOLO cao nhất cho tới thời điểm hiện tại là YOLO V6-V8. Mô hình sinh viên nhận chuyển giao tri thức sử dụng là YOLO V3 tiny. Kết quả đánh giá và so sánh hiệu quả của quá trình chuyển giao tri thức giữa các mô hình YOLO giáo viên khác nhau với cùng một mô hình sinh viên được trình bày chi tiết trong phần 3.

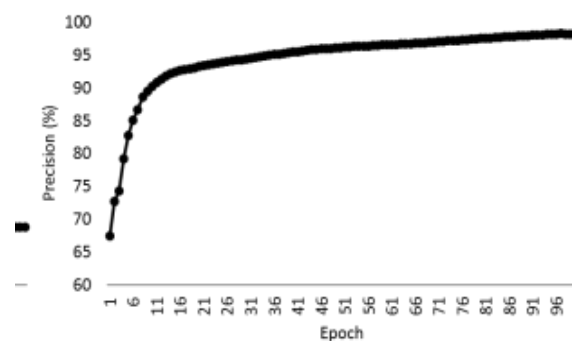
CSDL sử dụng là các bộ CSDL về vũ khí trong mục 2.2. Trước hết, ba mô hình YOLO phiên bản mới nhất là YOLO V6, YOLO V7 và YOLO V8 lần lượt được sử dụng làm mô hình chuyên gia để huấn luyện (retrain) và tinh chỉnh lại tham số (finetune) mô hình trên CSDL người và vũ khí do tác giả tự thu thập và gán nhãn. Sau đó, mỗi mô hình đã tinh chỉnh tham số này được đóng băng các lớp mạng và sử dụng với vai trò là giáo viên (teacher). Mô hình YOLO V3-tiny với vai trò là sinh viên (student) được thiết kế song song với mô hình giáo viên khi huấn luyện lại mô hình sinh viên nhằm mục đích kế thừa tri thức mà mô hình chuyên gia đã học được trước đó. Tham số khi thực hiện huấn luyện mô hình được thể hiện như trong Bảng 1.

Bảng 1. Các tham số triển khai quá trình huấn luyện hệ thống

Mô hình giáo viên (Teacher)	YOLO V6; YOLO V7; YOLO V8
Mô hình sinh viên (Student)	YOLO V3 tiny
Tốc độ huấn luyện (Learning rate)	10^{-5}
Số lượng vòng lặp (Epoch)	100
Số ảnh một mẻ (Batch size)	32 images
Kích thước ảnh	640×480 pixels



(a) Giá trị mất mát (Loss value)



(b) Giá trị độ chính xác Precision

Hình 2. Quá trình huấn luyện mô hình

Hình 2 minh họa kết quả quá trình huấn luyện mô hình YOLO V3 tiny. Kết quả này cho thấy hệ thống đạt hiệu suất tốt và bão hòa tại vòng lặp 90. Do vậy, trong các thử nghiệm trong mục 3 sẽ sử dụng mô hình này để thực hiện các đánh giá khác trong hệ thống đề xuất.

2.2.2. Phát hiện người và vũ khí

Mô hình YOLO V3 tiny sau khi được huấn luyện trong phần 2.1.1 sẽ được sử dụng để phát hiện vùng người và phát hiện vùng vũ khí (Hình 1b). Khi một ảnh I_{RGB} cho qua mô hình YOLO sẽ thu được các vùng ảnh người ($I_B^i (i = (1, \dots, M))$) và các vùng ảnh vũ khí ($I_W^j (j = (1, \dots, N))$) (Hình 3). Trong đó, M là số lượng người phát hiện được trong một ảnh, N là số lượng vũ khí phát hiện được trong một ảnh. Tất cả vùng người và các vùng vũ khí đã phát hiện được trong mỗi ảnh sẽ được cho vào bộ đệm S_1 và S_2 như mô tả trong công thức (3) và (4). Hai bộ đệm này được sử dụng để làm đầu vào cho quá trình phát hiện hành vi mang vũ khí ở phần 2.2.3.

$$S_1 = M^{YOLO}(I_{RGB}) = \{I_B^i, i = (1, \dots, M)\} \quad (3)$$

$$S_2 = M^{YOLO}(I_{RGB}) = \{I_W^j, j = (1, \dots, N)\} \quad (4)$$

Việc đánh giá định lượng hiệu quả của mô hình YOLO V3 tiny khi học chuyển giao từ các mô hình giáo viên khác nhau này sẽ được thực hiện và trình bày chi tiết trong mục 3.2.1.

2.2.3. Phát hiện hành vi mang vũ khí

Hệ thống phát hiện hành vi mang vũ khí

của người được đề xuất như mô tả trong Hình 1c nhằm tránh các trường hợp phát hiện sai. Hành vi mang vũ khí có thể là người cầm vũ khí trên tay hoặc vác, đội, đeo,... vũ khí trên người. Sau khi phát hiện các vùng người và các vùng vũ khí có trong một ảnh, chúng tôi sẽ duyệt các phần giao nhau giữa người và vũ khí trong hai bộ đệm S_1 và S_2 . Nếu trong các vùng người trong S_1 có sự giao nhau với vùng vũ khí trong S_2 thì vùng vũ khí đó được giữ lại bộ đệm như mô tả trong công thức (5) sau đây:

$$(I_W^k, I_B^j) = \begin{cases} (I_W^j, I_B^i) & I_B^i \cap I_W^j \neq \emptyset \\ remove & other \end{cases} \quad (5)$$

Trong đó, $j=(1,\dots,N)$, $k=(1,\dots,K)$, $K \leq N$. Sự dịch chuyển quang học (Optical Flow) của cả vùng vũ khí (I_W^k) và vùng người (I_B^j) sẽ được tính toán sử dụng mô hình RAFT (Recurrent All-Pairs Field Transforms for Optical Flow) [21] giữa khung hình hiện tại và khung hình trước đó lần lượt là $OF_{I_W^k}$ và $OF_{I_B^j}$. Trong đó, mỗi vùng người và/hoặc vũ khí với kích thước (P,Q) sẽ được sử dụng để tính vector luồng dịch chuyển quang học. Mỗi vùng được chia lưới đều nhau với kích thước lưới là 10×10 pixels. Như vậy, mỗi vùng sẽ có số lượng vector thể hiện luồng dịch chuyển quang học (H) là phần nguyên của $(P \times Q)/100$ và mỗi điểm trên lưới sẽ có giá trị thể hiện độ lớn của luồng dịch chuyển quang học là $O_h (h = (0, \dots, H))$. Sau đó, để đại diện cho sự dịch chuyển của mỗi vùng người

và/hoặc vũ khí được tính như trong công thức sau đây:

$$OF_{I_W^k} \frac{\sum_{h=1}^H O_h^{I_W^k}}{H} \text{ hoặc } OF_{I_B^k} \frac{\sum_{h=1}^H O_h^{I_B^k}}{H} \quad (6)$$

Hành vi mang vũ khí được lọc lại thông tin trước khi nhân cuối cùng được quyết định nhằm loại trừ trường hợp nhận sai như trong công thức (7):

$$I_W^k = \begin{cases} False & OF_{I_W^k} > thres \text{ and } OF_{I_B^k} < thres \\ True & others \end{cases} \quad (7)$$

Công thức này kiểm tra nếu vùng người dịch chuyển trong khi vùng vũ khí đứng yên thì vũ khí phát hiện được không thuộc hành vi mang vũ khí. Và nó chỉ được coi là hành vi mang vũ khí trong các trường hợp sau đây:

- Trường hợp 1 là vũ khí nằm gần người và cả vũ khí và người đều dịch chuyển khi người dịch chuyển.
- Trường hợp 2 là vũ khí nằm gần người và cả vũ khí và người cùng đứng yên.
- Trường hợp 3 là tay cầm vũ khí và thay đổi nhưng người vẫn đứng yên.

Việc lọc lại hành vi như trên nhằm loại bỏ các trường hợp có vũ khí xuất hiện trong khung hình nhưng người không sử dụng vũ khí đó thì không được coi là hành vi mang vũ khí. Với giải pháp đề xuất đã khắc phục được trong điều kiện nền phức tạp có vật thể giống vũ khí hoặc trong nền có xuất hiện vũ khí và người hoàn toàn không sử dụng vũ khí đó.

2.2. Cơ sở dữ liệu vũ khí WeaponEpuMq

Các bộ CSDL vũ khí hiện có do cộng đồng nghiên cứu cung cấp miễn phí như Weapon [15]-[20] thường cung cấp số lượng nhân vũ khí khá giới hạn với chỉ từ hai đến bốn loại vũ khí. Ngoài ra, các CSDL vũ khí này thường là các vùng ảnh với nền đơn giản và chỉ có vũ khí ở những hình trạng khá đơn giản, rõ nét và chính diện. Trong nghiên cứu này chúng tôi thu thập CSDL có hành vi người mang vũ khí. Trong đó, chúng tôi sử dụng 06 loại vũ khí như minh họa trong Hình 3.



(a) Dao

(b) Súng



(c) Kiếm/ dao



(d) Đinh ba



(e) Phóng lợn/ mã tấu



(f) Côn

Hình 3. Minh họa các loại vũ khí trong bộ CSDL WeaponEpuMq

Chúng tôi đặt tên bộ CSDL vũ khí tự thu thập với tên là WeaponEpuMq. Đặc điểm của bộ CSDL này là có người xuất hiện trong các khung hình. Trong đó, có một hoặc nhiều người mang vũ khí và có người không mang vũ khí. Ngoài ra, trong khung hình xuất hiện các vũ khí nhưng không được người mang/cầm. Bộ CSDL này được thu thập tại các ngữ cảnh trong nhà và ngoài trời và tại các thời điểm hoàn toàn khác nhau trong nhiều ngày. CSDL WeaponEpuMq có tất cả 100 video. Máy quay sử dụng là HiK-Vision DS-2CD2643G2-IZS cung cấp ảnh màu RGB với tốc độ thu thập 30 fps, độ phân giải ảnh 2688×1520 pixels. Mỗi video có độ dài 1 phút đến 10 phút. Tổng số lượng nhãn các loại người, vũ khí và hành vi mang vũ khí trong bộ CSDL này được trình bày chi tiết như trong Bảng 2.

Bảng 2. Số lượng từng loại dữ liệu trong CSDL WeaponEpuMq

STT	Loại vũ khí	Số lượng
1	Dao	26662
2	Súng	28518
3	Kiểm/Đao	25567
4	Đinh ba	26781
5	Phóng lựu	24193
6	Côn	27291
7	Người	35671
8	Hành vi mang vũ khí	125890

Bộ CSDL này được chúng tôi gán nhãn thành hai phần: (1) Loại gán nhãn thứ nhất thực hiện gán cho từng vùng hình

ảnh có chứa 06 loại vũ khí trong Hình 3 và gán nhãn vùng hình ảnh chứa người; (2) Loại gán nhãn thứ hai chúng tôi thực hiện gán nhãn hành vi mang vũ khí của người là vùng hình ảnh mà người mang/đeo/cầm một trong sáu loại vũ khí trên. Hình 4 minh họa một số hình ảnh trong bộ CSDL của chúng tôi tự thu thập tại các điều kiện nền và ngữ cảnh khác nhau. Hình 4a minh họa hình ảnh thu thập tại môi trường ngoài trời, Hình 4b minh họa hình ảnh thu thập tại môi trường trong nhà.



(a) Thu thập ngoài trời



(b) Thu thập trong nhà

Hình 4. Minh họa một số hình ảnh của bộ CSDL WeaponEpuMq

2.3. Giao thức và thang đo đánh giá

2.3.1. Thang đo đánh giá (Metric)

Để đánh giá hiệu quả của mô hình YOLO,

chúng tôi sử dụng các chỉ tiêu độ chính xác P (Precision), độ triệu hồi R (Recall). Trong đó, độ chính xác thu được trên tập dữ liệu thử nghiệm là thước đo đánh giá độ hiệu quả của mô hình qua các thông số độ chính xác P và độ triệu hồi R. Độ chính xác P được định nghĩa là tỉ lệ của số người phát hiện đúng trên tổng số nhãn được phát hiện minh họa bởi công thức sau đây:

$$P = \frac{TP}{TP+FP} \quad (8)$$

Độ triệu hồi R được định nghĩa là tỉ lệ của số nhãn phát hiện đúng trên tổng số nhãn có trong ảnh như minh họa bởi công thức sau đây:

$$R = \frac{TP}{TP+FN} \quad (9)$$

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

2.3.2. Giao thức (Protocol)

Trong bài báo này, chúng tôi chia bộ CSDL thành 10 phần bằng nhau. Sau đó, chúng tôi sử dụng phương pháp “*Leave-one-subject-out*” [23] để thực hiện 10 lần thử nghiệm. Trong đó, tại mỗi lần thử nghiệm, tám phần dữ liệu được sử dụng để huấn luyện mô hình, một phần để

validation và một phần để đánh giá (test) mô hình. Tại mỗi lần thử nghiệm này sẽ có độ chính xác P_j , R_j ($j=(1-M)$, $M=10$). Độ chính xác của một CSDL đạt được bằng cách lấy trung bình của 10 lần đánh giá mô hình với độ chính xác P (Precision) như trong công thức (10) sau:

$$P = \frac{\sum_{j=1}^{M=10} P_j}{M} \quad (10)$$

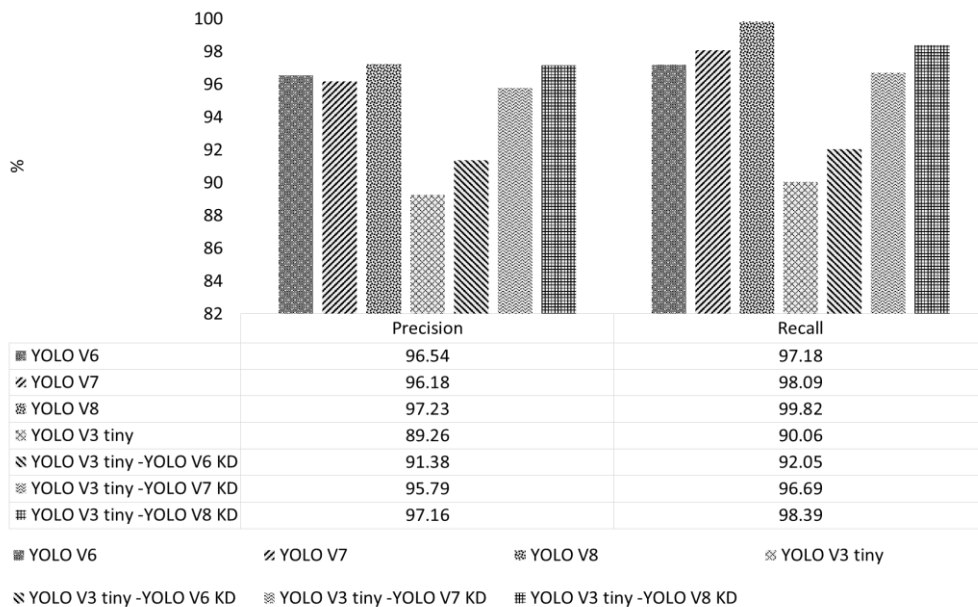
Độ triệu hồi R (recall) của cả cơ sở dữ liệu được tính như trong công thức (11) sau:

$$R = \frac{\sum_{j=1}^{M=10} R_j}{M} \quad (11)$$

3. KẾT QUẢ THỬ NGHIỆM

Chúng tôi thực hiện các thử nghiệm trên máy tính sử dụng CPU Intel Core i5-11400H, GPU NVIDIA GeForce GTX 1650, bộ nhớ RAM 8 GB. Ngôn ngữ lập trình sử dụng là ngôn ngữ Python. Các thử nghiệm được thực hiện trên 01 bộ CSDL dùng chung Pistol detection dataset [18] và 01 bộ CSDL tự thu thập có tên là WeaponEpuMq. Trong đó, bộ CSDL Pistol detection dataset [18] có 3000 ảnh của các loại súng ngắn.

Các thử nghiệm được tiến hành trong nghiên cứu này gồm: (1) Đánh giá độ chính xác phát hiện người và vũ khí với các mô hình KD từ các mô hình chuyên gia/giáo viên khác nhau; (2) Đánh giá độ chính xác của mô hình phát hiện hành vi mang vũ khí của người; (3) Đánh giá thời gian đáp ứng của mô hình (Time cost) học chuyển giao và thời gian đáp ứng của hệ thống phát hiện hành vi mang vũ khí.



Hình 5. Độ chính xác (P-Precision), độ triệu hồi (R-Recall) cho phát hiện vũ khí trên CSDL WeaponEpuMq

3.1. So sánh kết quả quá trình học chuyển giao (KD)

Trong phần này, chúng tôi sẽ tiến hành khảo sát sự ảnh hưởng của việc sử dụng các mô hình giáo viên khác nhau khi chuyển giao tri thức cho cùng một mô hình sinh viên (Hình 1a).

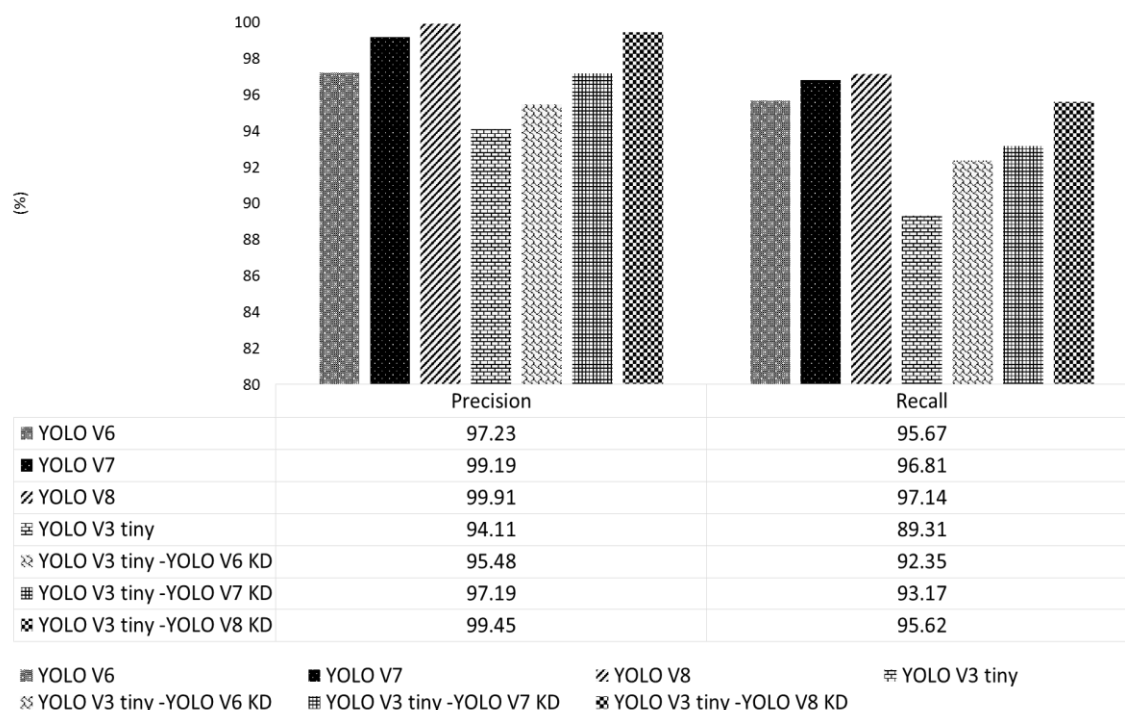
Trong đó, ba mô hình giáo viên lần lượt là YOLO V6, YOLO V7 và YOLO V8. Mỗi thời điểm chỉ sử dụng một mô hình để chuyển giao tri thức cho mô hình YOLO tiny V3. Hình 5 và Hình 6 biểu diễn kết quả của quá trình phát hiện vùng người và vùng vũ khí của các mô hình so với kết quả mà chúng tôi đã gắn nhãn.

Hình 5 biểu diễn kết quả phát hiện các loại vũ khí trên bộ CSDL tự thu thập WeaponEpuMq với IoU = 0,5, kết quả trong cho thấy một số điểm nổi bật như sau:

- Trong ba mô hình chuyên gia lần lượt từ YOLO V6-V8 thì việc sử dụng mô hình YOLO V8 để phát hiện đạt kết quả cao nhất trên hai tiêu chí đánh giá P và R với 97,23% và 98,82%. Tuy nhiên, ba mô hình này có cấu trúc hình phức tạp nên khi chạy trực tuyến yêu cầu cấu hình máy cao và thời gian đáp ứng lâu (trong mục 3.3.3).
- Nếu sử dụng mô hình YOLO V3 tiny có yêu cầu cấu hình máy nhẹ hơn và thời gian đáp ứng nhanh hơn so với ba phiên bản đầy đủ của YOLO V6-8. Tuy nhiên, mô hình này lại đạt kết quả khá thấp trên cả thang đo độ triệu hồi (P), độ chính xác (R) khi phát hiện vùng người và vùng vũ khí tại các giá trị tương ứng là 89,26% và 90,06%.
- Khi sử dụng một trong số các mô hình YOLO V6-8 để huấn luyện lại và tính

chính tham số cho YOLO V3 tiny thì kết quả của YOLO V3 tiny KD đạt kết quả tốt hơn trên cả ba loại học chuyển giao. Trong đó, với mô hình học chuyển giao mà giáo viên là YOLO V8 thì kết quả của YOLO V3 tiny đạt được kết quả cao nhất trên cả ba tiêu chí đánh giá (Hình 5) và cao hơn hẳn so với mô hình YOLO V3 tiny (Hình 5). Kết quả này cho thấy hiệu quả của quá trình học chuyển giao có chất lượng đạt hiệu quả cao hơn đối với mô hình YOLO V3 tiny khi áp dụng cho bài toán phát hiện vũ khí. Do đó, mô hình YOLO V3 tiny - YOLO V8 KD sẽ được sử dụng trong đánh giá ở các mục 3.2.2 và 3.2.3. Kết quả ma trận nhầm lẫn chi tiết của phát hiện vũ khí cho hai giải pháp sử dụng mô hình YOLO V3 tiny và YOLO

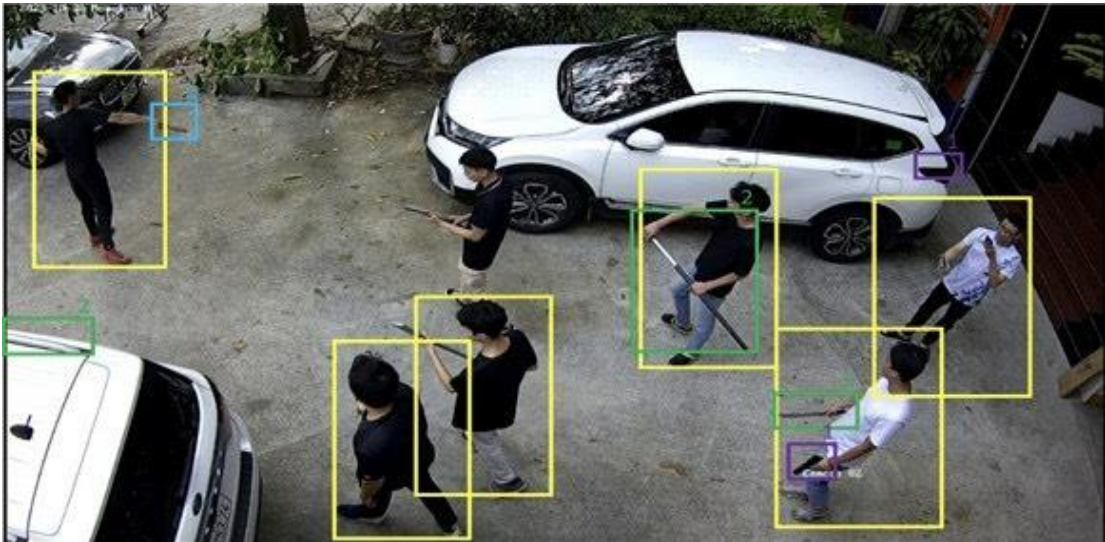
V3 tiny - YOLO V8 KD được thể hiện như trong Hình 8. Trong đó, phương pháp YOLO V3 tiny kết quả đạt được thấp hơn hẳn trên cả sáu lớp dữ liệu so với phương pháp YOLO V3 tiny - YOLO V8 KD. Đặc biệt là phát hiện thiếu, phát hiện nhầm sang nền hoặc sang loại vũ khí khác. Trong đánh giá này, tác giả cũng thực hiện tính toán độ chính xác phát hiện người trong ảnh của các phương pháp khác nhau trên bộ CSDL WeaponEpuMq tại IoU = 0,5 như mô tả trong Hình 6. Kết quả này một lần nữa cho thấy việc áp dụng mô hình học chuyển giao từ hệ chuyên gia (YOLO V6-V8) đem đến kết quả tốt hơn cho phát hiện người khi sử dụng mô hình YOLO V3 tiny. Kết quả cải thiện trên cả hai tiêu chí Precision và



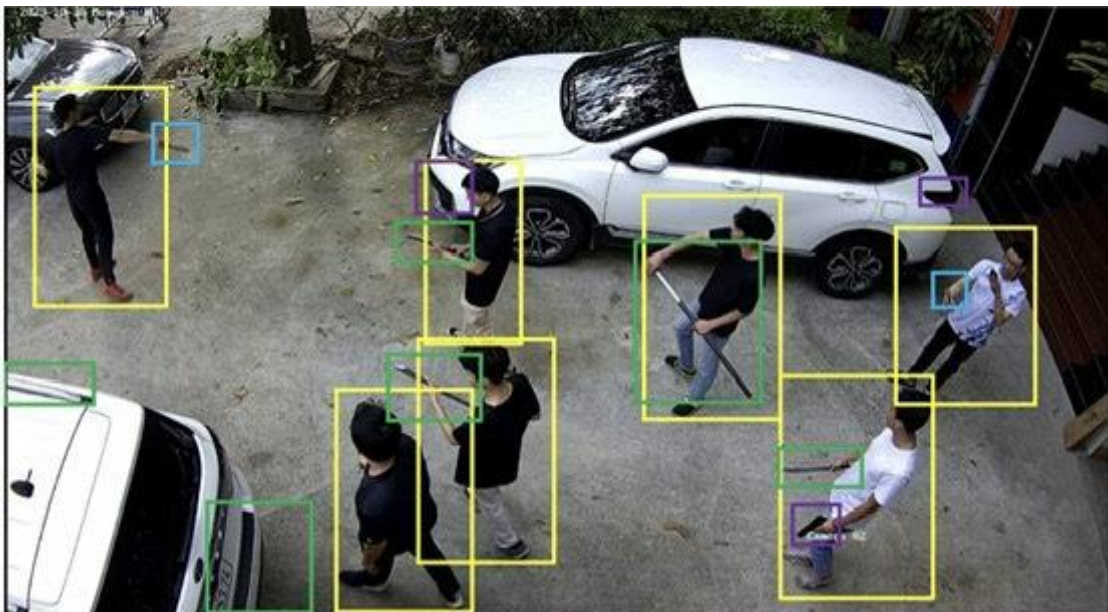
Hình 6. Độ chính xác (P-Precision), độ triệu hồi (R-Recall) cho phát hiện người trên CSDL WeaponEpuMq

Recall và kết quả phát hiện vùng người tốt nhất trên mô hình học chuyển giao YOLO V3 tiny - YOLO V8 KD với Precision là 99,45% và Recall là 95,62%. Trong các đánh giá sau chúng tôi sẽ sử dụng mô hình này để thực hiện các thử nghiệm.

Hình 7 minh họa kết quả của quá trình phát hiện vùng người và vùng vũ khí trên ảnh sử dụng với hai phiên bản YOLO V3 tiny (Hình 7a và YOLO V3 tiny - YOLO V8 KD (Hình 7b).



a) Kết quả của mô hình YOLO V3 tiny



b) Kết quả của mô hình YOLO V3 tiny –YOLO V8KD

Hình 7. Kết quả phát hiện vùng người và vùng vũ khí

Trong đó, Hình 7a vẫn còn một vùng người và vùng vũ khí không phát hiện được khi sử dụng mô hình YOLO V3 tiny và vùng này được phát hiện khi sử dụng mô hình học chuyển giao của YOLO V3 tiny - YOLO V8 KD (Hình 7b).

Tuy nhiên, khi sử dụng mô hình YOLO V3 tiny - YOLO V8 KD lại phát hiện nhầm vùng nền có đối tượng hình ảnh giống vũ khí đã được gán nhãn nằm trên ô tô (hai vùng có đường bao ô vuông màu xanh lá cây trên ô tô phát hiện nhầm là kiếm và hai vùng bao màu tím phát hiện nhầm thành súng trong Hình 7b). Hình 7 trên phát hiện tất cả 07 người trong ảnh là vùng có khung màu vàng và hệ thống phát hiện ra 11 vũ khí trong ảnh là các vùng có khung với các màu xanh lá (kiếm), xanh dương (dao), tím (súng). Nếu chỉ sử dụng mô hình YOLO để phát hiện vũ khí trong ảnh thì sẽ bị nhận nhầm khi trong ảnh có chứa vũ khí nhưng người không sử dụng chúng. Hoặc hệ thống sẽ nhận nhầm khi trong nền có vùng ảnh có hình trạng giống vũ khí. Ngoài ra, trong nghiên cứu này chúng tôi cũng đánh giá trên CSDL dùng chung Pistol detection dataset [18] với các phương pháp sử dụng như trình bày trong Bảng 3. Kết quả cho thấy việc áp dụng mô hình chuyển giao tri thức mô hình YOLO V8 cho mô hình YOLO V3 tiny (hàng 4 của Bảng 3) cũng đạt hiệu quả cao hơn so với sử dụng mô hình YOLO V3 tiny (hàng 2 của Bảng 3). Hơn nữa, kết quả đạt được cũng cao hơn giải

pháp trong [18] sử dụng mô hình fast R-CNN với backbone là VGG16 (hàng 1 của Bảng 3).

Bảng 3. Kết quả phát hiện vũ khí của CSDL Pistol detection dataset [18]

Phương pháp	Precision	Recall
[18]	89,91	32,24
YOLO V3 tiny	83,12	31,89
YOLO V8	98,51	97,19
YOLO V3 tiny - YOLO V8 KD	97,36	94,74

Giải pháp đề xuất để phát hiện hành vi mang vũ khí của người được cải thiện so với việc sử dụng mô hình YOLO và kết quả được thể hiện chi tiết trong mục 3.2.

3.2. Hiệu quả của mô hình phát hiện hành vi mang vũ khí

Trong đánh giá này chúng tôi sử dụng kết quả vùng người và vùng vũ khí đã phát hiện được từ mô hình YOLO V3 tiny - YOLO V8 KD như đã trình bày trong mục 3.2.1. Đây là mô hình được lựa chọn vì đạt hiệu quả cao hơn. Từ hai luồng thông tin là vùng người và vùng vũ khí đã phát hiện được, hành vi mang vũ khí được xác định dựa trên giải pháp đề xuất như đã trình bày trong mục 2.2.3. Trong việc thực hiện gán nhãn cho CSDL vũ khí của chúng tôi, ngoài việc gán nhãn vùng người và vùng vũ khí thì chúng tôi còn gán nhãn hành vi mang vũ khí của người trong khung hình. Do vậy, trong mục này, chúng tôi thực hiện so sánh kết quả của việc phát hiện hành vi mang vũ khí đã

	Đao	Súng	Kiểm	Đao	Mã tấu	Còn		Đao	Súng	Kiểm	Đao	Mã tấu	Còn		Đao	Súng	Kiểm	Đao	Mã tấu	Còn
Đao	0.811	0.030	0.090	0.010			Đao	0.823		0.008	0.002			Đao	0.946		0.008	0.002		
Súng		0.832					Súng		0.993					Súng		0.958				
Kiểm	0.040		0.813	0.040	0.050	0.017	Kiểm	0.009		0.915	0.006	0.017		Kiểm	0.009		0.968	0.006	0.017	
Đao		0.050	0.023	0.814	0.010		Đao			0.008	0.926	0.006		Đao			0.008	0.978	0.006	
Mã tấu	0.064		0.063	0.020	0.906		Mã tấu			0.009	0.004	0.984		Mã tấu			0.009	0.004	0.976	
Còn						0.912	Còn					0.958		Còn						0.980
Nền	0.085	0.088	0.011	0.116	0.034	0.071	Nền	0.068	0.069	0.060	0.062	0.043	0.042	Nền	0.045	0.042	0.007	0.010	0.001	0.011

a) b) c)

Hình 8. Bảng ma trận nhầm lẫn (Confusion Matrix) phát hiện hành vi mang vũ khí

được gán nhãn với hành vi mang vũ khí mà hệ thống phát hiện được. Trong đó, độ chính xác của việc phát hiện hành vi mang vũ khí được tính dựa trên các tiêu chí như đã trình bày trong mục 2.3. Kết quả được thể hiện như trong Bảng 4.

Bảng 4. Kết quả phát hiện hành vi mang vũ khí của CSDL WeaponEpuMq

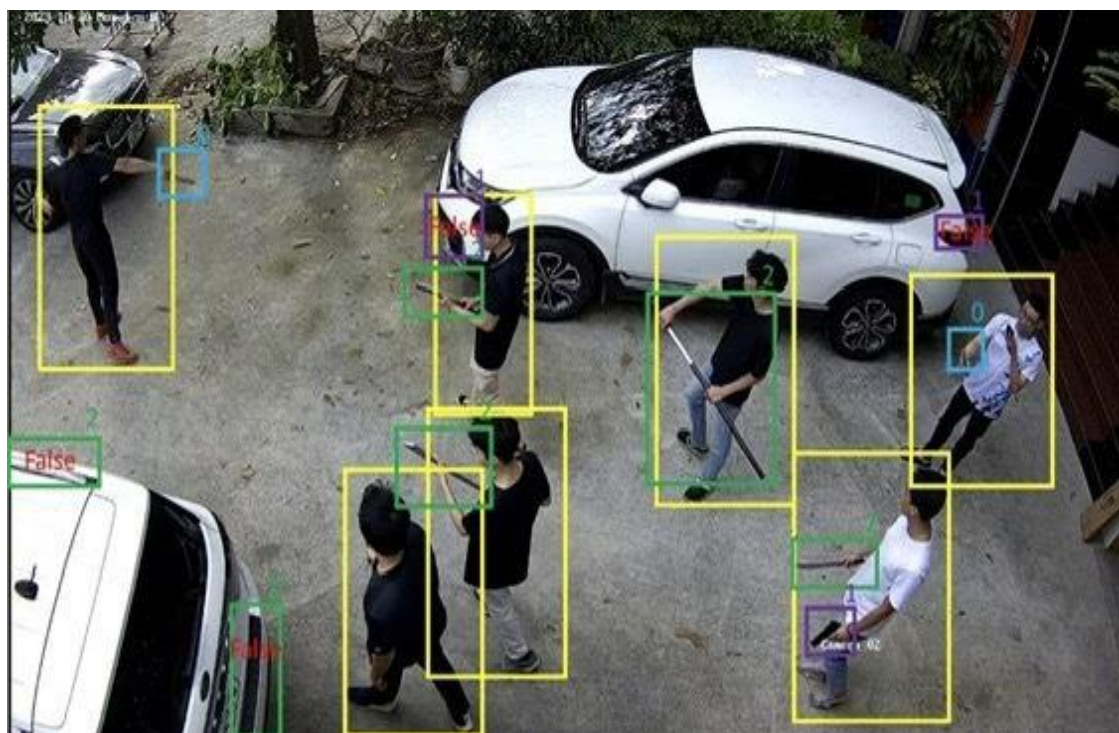
Phương pháp	Precision	Recall
YOLO V3 tiny	84,80	82,06
YOLO V3 tiny - YOLO V8 KD	93,12	91,39
Proposed method	96,2	95,94

Kết quả trong Bảng 4 cho thấy giải pháp đề xuất cải thiện đáng kể việc nhận nhầm/sai so với việc chỉ phát hiện vùng vũ khí là kết quả đầu ra của YOLO V3

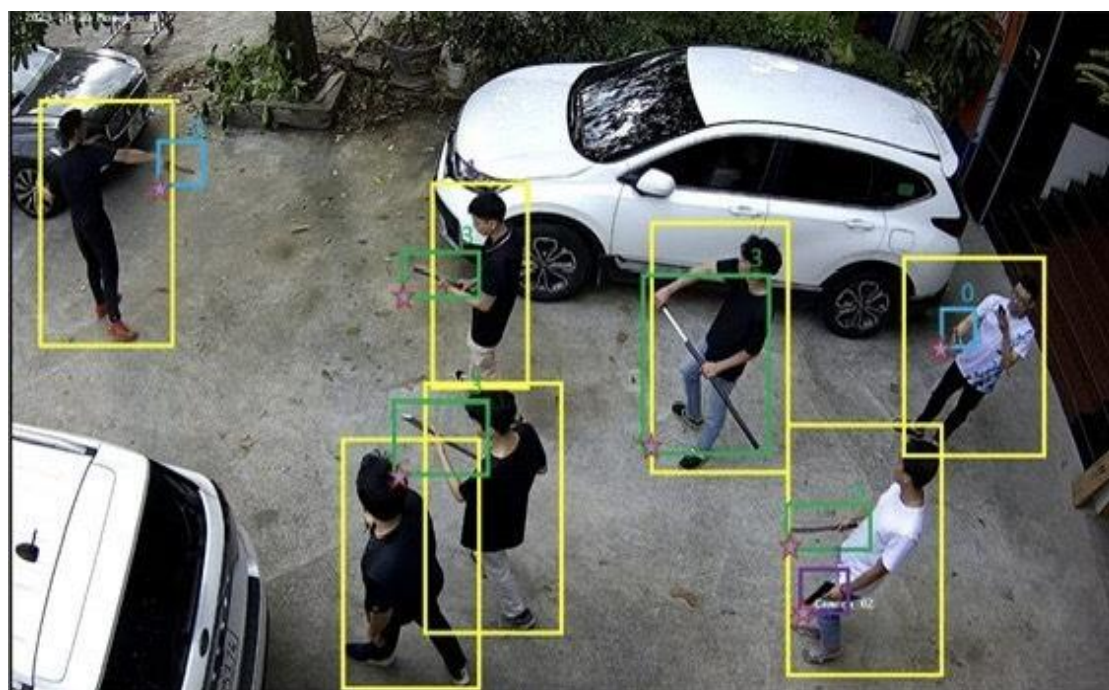
tiny và YOLO V3 tiny - KD. Trong đó, P tăng từ 84,8% lên 96,9%, R tăng từ 82,06% lên 95,94%. Hình 8 biểu diễn ma

trận nhầm lẫn của ba phương pháp YOLO V3 tiny; YOLO V3 tiny - YOLO V8 KD; YOLO V3 tiny - YOLO V8 KD với luồng dịch chuyển quang học của vùng người và vùng vũ khí (mục 2.2.3). Kết quả cho thấy hệ thống phát hiện hành vi mang vũ khí đề xuất nâng cao độ chính xác phát hiện nhầm so với sử dụng đầu ra của mô hình YOLO. Đặc biệt giảm đáng kể trường hợp phát hiện nhầm sang đối tượng nền có hình trạng giống vũ khí.

Hình 9 minh họa kết quả phát hiện hành vi mang vũ khí. Kết quả cho thấy việc sử dụng mô hình YOLO phát hiện nhầm 04 đối tượng nền (Hình 9a). Tuy nhiên, khi sử dụng giải pháp của tác giả đề xuất thì đã loại trừ được các trường hợp phát hiện nhầm (Hình 9b).



a) Kết quả của mô hình YOLO V3 tiny-YOLO V8 KD



b) Kết quả của mô hình YOLO V3 tiny-YOLO V8 KD

Hình 9. Kết quả phát hiện hành vi mang vũ khí

3.3. Thời gian đáp ứng của mô hình KD

Ngoài việc so sánh kết quả độ chính xác R, độ triệu hồi P đã trình bày ở các phần trên. Trong mục này, chúng tôi tiến hành thực hiện khảo sát về thời gian thực hiện trung bình của các phiên bản YOLO cho hai quá trình: (1) Quá trình phát hiện người và vũ khí; (2) Quá trình phát hiện hành vi sử dụng vũ khí. Đánh giá về thời gian đáp ứng này được chạy thử nghiệm trên nền tảng CPU và GPU. Kết quả thời gian đáp ứng được trình bày trong Bảng 5.

Bảng 5. Tốc độ phát hiện người và vũ khí trên các phiên bản YOLO

Phương pháp	Tốc độ xử lý một khung hình sử dụng CPU (ms)	Tốc độ xử lý một khung hình sử dụng GPU (ms)
YOLO V6	80,39	22,09
YOLO V7	85,03	23,81
YOLO V8	78,85	21,74
YOLO V3 tiny	21,05	12,02
YOLO V3 tiny -YOLO V8 KD	20,98	11,86

Kết quả trong Bảng 5 cho thấy mạng YOLO V3 tiny KD có thời gian phát hiện khi sử dụng GPU là thấp nhất và tương đương mô hình YOLO V3 tiny với 12 ms và khi sử dụng CPU thì kết quả là 21 ms. Kết quả này thấp hơn rất nhiều so với việc sử dụng các mô hình YOLO V6 đến V8 có thời gian đáp ứng khi sử dụng CPU lần lượt là 80,39 ms, 85,03 ms và 78,85 ms.

Thời gian đáp ứng khi sử dụng GPU lần lượt là 22,09 ms, 23,81 ms và 21,74 ms. Trong khi đó, độ chính xác phát hiện của mô hình YOLO V3 tiny KD có độ chính xác cao hơn YOLO V3 tiny và độ chính xác tiệm cận với kết quả của các mô hình phức tạp là YOLO V7 và YOLO V8. Điều này cho thấy giải pháp học chuyển giao có chất lọc tri thức với mạng YOLO do chúng tôi đề xuất đạt kết quả tốt với mô hình phát hiện và người và vũ khí trong thử nghiệm trên CSDL WeaponEpuMq.

Ngoài ra, chúng tôi cũng thực hiện đánh giá thời gian trung bình để hệ thống phát hiện hành vi người mang vũ khí khi sử dụng mô hình YOLO V3 tiny - YOLO V6 KD từ thời điểm đọc ảnh đến khi xử lý xong một khung hình. Trong đó, thời gian thực hiện trên CPU là 26,89 ms và thời gian thực hiện trên GPU là 14,59 ms.

4. THẢO LUẬN VÀ KẾT LUẬN

Trong bài báo này, chúng tôi đã thực hiện học chuyển giao có chất lọc tri thức từ mô hình giáo viên cho mạng YOLO và áp dụng cho CSDL phát hiện vũ khí. Kết quả cho thấy mô hình học chuyển giao khá hiệu quả nhằm đạt được một mô hình có kiến trúc nhẹ hơn, thời gian đáp ứng nhanh hơn. Các thử nghiệm được thực hiện trên bộ CSDL do chúng tôi tự thu thập WeaponEpuMq và bộ CSDL dùng chung. Kết quả cho thấy giải pháp hiệu chỉnh tham số có kế thừa tri thức từ mạng

chuyên gia đạt hiệu quả cao hơn từ 1,99% đến 7,9% trên cả hai bộ CSDL, trong khi thời gian đáp ứng của mô hình giảm 62 ms khi sử dụng CPU và 10,95 ms khi sử dụng GPU. Nghiên cứu còn đề xuất giải pháp phát hiện hành vi mang vũ khí của người với việc sử dụng kết hợp vùng người và vùng vũ khí đã phát hiện được. Ngoài ra, thông tin về luồng dịch chuyển quang học của hai vùng phát hiện được này cũng được sử dụng để phát hiện hành vi mang vũ khí của người. Các đánh giá thử nghiệm trên bộ CSDL do chúng tôi thu thập về hành vi mang vũ khí cho thấy

giải pháp đề xuất đạt kết quả tốt hơn so với việc chỉ sử dụng kết quả phát hiện vũ khí từ đầu ra của mạng YOLO V3 tiny tới 13,88%. Trong thời gian tới, ý tưởng đề xuất của nghiên cứu này sẽ tiếp tục được triển khai để đánh giá trên bộ CSDL lớn và đa dạng hơn cũng như áp dụng cho phát hiện hành vi mang các vật dụng khác của người. Việc khảo sát hiệu quả của các phương pháp tính toán luồng quang học cũng sẽ được khảo sát và kết hợp với theo vết đối tượng để nâng cao hơn nữa hiệu quả của giải pháp phát hiện hành vi mang vũ khí.

TÀI LIỆU THAM KHẢO

- [1] 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," The fourth International Conference on Multimedia Analysis and Pattern Recognition (MAPR), HaNoi, VietNam, 2021.
- [2] Shah, S., Ghomeshi, H., Vakaj, E. et al. A review of natural language processing in contact centre automation. *Pattern Anal Applic* 26, 823–846 (2023).
- [3] Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali, 2016. *YOLOv1: You Only Look Once: Unified, Real-Time Object Detection*. 779-788. 10.1109/CVPR.2016.91.
- [4] Redmon Joseph, Farhadi Ali, 2017. *YOLO9000: Better, Faster, Stronger*. 6517-6525. 10.1109/CVPR.2017.690.
- [5] J. Redmon, A. Farhadi, 2018. *Yolov3: An incremental improvement*. CoRR journal, Vol. abs/1804.02767, pp. 1-6.
- [6] A. Bochkovskiy, C.Y. Wang, H.Y. M. Liao, 2004. *Yolov4: Optimal speed and accuracy of object detection*. ArXiv journal, Vol. abs/2004.10934, 2020, pp. 1-17.
- [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, 2022. *YOLOv6: A Single-Stage Object Detection*

Framework for Industrial Applications. Computer Vision and Pattern Recognition.

- [9] Wang Chien-Yao, Bochkovskiy Alexey, Liao Hong-yuan, 2022. *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.* 10.48550/arXiv.2207.02696.
- [10] Jocher, G.; Chaurasia, A.; Qiu, J. "YOLO by Ultralytics,". GitHub. 1 January 2023. Available online: <https://github.com/ultralytics/ultralytics>.
- [11] He, K., Zhang, X., Ren, S., Sun, J.: "Deep residual learning for image recognition," Computer Vision and Pattern Recognition (CVPR 2016) (2016).
- [12] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.
- [13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear Bottlenecks," 2018.
- [14] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," ArXiv, vol. abs/1503.02531, 2015.
- [15] Olmos, R., Tabik, S., & Herrera, F. (2018) *Automatic handgun detection alarm in videos using deep learning.* Neurocomputing, 275, 66-72 doi.org/10.1016/j.neucom.2017.05.012.
- [16] Castillo, A., Tabik, S., Pérez, F., Olmos, R., Herrera, F. (2019) *Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos using deep learning.* Neurocomputing, 330, 151-161. doi.org/10.1016/j.neucom.2018.10.076.
- [17] Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H., Herrera, F. (2020) *Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance.* Knowledge-Based Systems, 194, 105590. doi.org/10.1016/j.knosys.2020.105590.
- [18] Olmos, R., Tabik, S., & Herrera, F. (2018) *Automatic handgun detection alarm in videos using deep learning.* Neurocomputing, 275, 66-72. doi.org/10.1016/j.neucom.2017.05.012.
- [19] Castillo, A., Tabik, S., Pérez, F., Olmos, R., Herrera, F. (2019) *Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos using deep learning.* Neurocomputing, 330, 151-161. doi.org/10.1016/j.neucom.2018.10.076.
- [20] Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H., Herrera, F. (2020) *Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance.* Knowledge-Based Systems, 194, 105590. doi.org/10.1016/j.knosys.2020.105590.
- [21] R. Y. Rubinstein, "Optimization of computer simulation models with rare events," European Journal of Operational Research, vol. 99, no. 1, pp. 89–112, 1997.

- [22] Teed, Z., Deng, J. (2020). "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow". In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. Lecture Notes in Computer Science(), vol 12347. Springer, Cham.
- [23] Huong-Giang Doan, Ngoc-Trung Nguyen, "New blender-based augmentation method with quantitative evaluation of CNNs for hand gesture recognition", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 30, No. 2, 2023, pp. 796-806.
- DOI: 10.11591/ijeecs.v30.i2.pp796-806.

Giới thiệu tác giả:

Tác giả Đoàn Thị Hương Giang tốt nghiệp đại học ngành đo lường và tin học công nghiệp năm 2003, nhận bằng Thạc sĩ ngành đo lường và các hệ thống điều khiển năm 2006 và bằng Tiến sĩ ngành điều khiển và tự động hóa năm 2008 tại Đại học Bách khoa Hà Nội. Hiện nay tác giả là giảng viên Khoa Điều khiển và Tự động hóa - Trường Đại học Điện lực.

Lĩnh vực nghiên cứu: Điều khiển và tự động hóa sử dụng thông tin hình ảnh, xử lý ảnh, học máy và học sâu...

