

MÔ HÌNH SÚNG MÁY PHÁT HIỆN VÀ BẮM SÁT MỤC TIÊU TỰ ĐỘNG THỜI GIAN THỰC DỰA TRÊN MẠNG NƠ RON TÍCH CHẬP

REAL-TIME AUTOMATIC TARGET DETECTION AND TRACKING MODEL
BASED ON CONVOLUTIONAL NEURAL NETWORK FOR MACHINE GUN SYSTEM

Đinh Tuấn Anh¹, Nguyễn Tiến Hoàng¹,
Hà Việt Anh¹, Vũ Hữu Thích¹, Phạm Văn Hùng^{1,*}

DOI: <http://doi.org/10.57001/huiv5804.2024.308>

TÓM TẮT

Trong bài báo này, mô hình mô phỏng súng máy phát hiện và bám sát mục tiêu tự động ứng dụng cho máy bay quân sự sử dụng công nghệ in 3D được phát triển. Đồng thời đề xuất thuật toán tự động phát hiện và bám mục tiêu sử dụng mạng nơ-ron tích chập, cụ thể là sử dụng mô hình YOLOv8 để tăng tốc độ và độ chính xác nhận dạng, vi điều khiển STM32 để điều khiển hệ thống chuyển động bám mục tiêu. Sau khi thực nghiệm cho thấy hệ thống nhận dạng mục tiêu với độ chính xác khá cao 97,86% và bám mục tiêu một cách nhanh chóng khi mục tiêu chuyển động với tốc độ nhỏ hơn 0,5m/s.

Từ khóa: *Bắt và bám đối tượng thời gian thực; súng máy tự động; YOLOv8*

ABSTRACT

In this paper, we construct a simulation model of an automatic target detection and tracking machine gun application for military aircraft using 3D printing technology. Simultaneously, we propose an algorithm for automatic target detection and tracking using a convolutional neural network model, specifically based on the YOLOv8 model, to enhance the speed and accuracy of target identification. We utilize an STM32 microcontroller to control the target tracking motion system. Experimental results demonstrate that the target recognition system achieves a high accuracy rate of 97.86% and effectively tracks moving targets with speeds below 0.5m/s.

Keywords: *Real-time target detection and tracking; machine gun; YOLOv8*

¹Trường Đại học Công nghiệp Hà Nội

*Email: phamvanhung@hauivn.edu.vn

Ngày nhận bài: 05/4/2024

Ngày nhận bài sửa sau phản biện: 05/6/2024

Ngày chấp nhận đăng: 27/9/2024

1. GIỚI THIỆU

Hiện nay, việc ứng dụng xử lý ảnh và AI (Artificial Intelligent) trong việc tự động nhận dạng và bắt bám đối

tượng ngày một trở nên phổ biến. Ứng dụng mạng nơ-ron để nhận dạng biển số xe được đề cập trong tài liệu [1] hay hệ thống báo cháy tự động được đề cập đến trong [2] hoặc phát hiện và theo dõi con người cho hệ thống chống trộm [3]. Còn trong lĩnh vực trong quân sự tài liệu [4-7] là các nghiên cứu về hệ thống thời gian thực sử dụng xử lý ảnh và AI để theo dõi mục tiêu và điều khiển vũ khí di động, giúp tăng hiệu quả chiến đấu và đảm bảo sự chính xác trong các tình huống chiến đấu. Những năm gần đây, các cường quốc trên thế giới đang đầu tư nhân lực, chất xám, tiền bạc vào phát triển những hệ thống hỏa lực tự động trên các phương tiện cơ giới mặt đất, máy bay chiến đấu, tàu hải quân... Trong số đó phải kể đến hệ thống điều khiển đài quang điện tử cỡ nhỏ tích hợp trên UAV do công ty CACI của Hoa Kỳ hay hệ thống Slinger điều khiển pháo M230LF do Công ty công nghệ quân sự EOS [8] của Australia nghiên cứu và sản xuất.

Ở Việt Nam một số đơn vị đang tích cực phát triển những hệ thống hỏa lực thông minh điều khiển tự động thông qua máy chính trung tâm. Một trong số đó là hệ thống súng máy tự động 12,7mm được Viện Vũ khí (Tổng cục Công nghiệp Quốc Phòng) nghiên cứu và chế tạo [9]. Ở quy mô lớn hơn, Tổng công ty Công nghiệp Công nghệ cao Viettel đã tiến hành nghiên cứu phát triển các Đài quang điện tử, gồm hệ thống phát hiện và hệ thống khai hỏa [10], ứng dụng cho hải quân, không quân và lục quân giúp phát hiện đối tượng dưới điều kiện thời tiết khác nhau, giúp tăng hiệu quả chiến đấu.

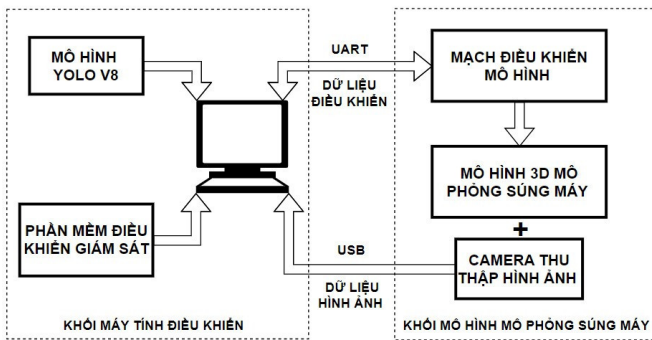
Như vậy có thể thấy nghiên cứu áp dụng công nghệ cao trong lĩnh vực quân sự hiện đang rất được thế giới quan tâm và phát triển, tuy nhiên ở Việt Nam các nghiên cứu vẫn còn khá hạn chế. Với mong muốn tiếp cận ứng dụng công nghệ cao cũng như làm chủ công nghệ, chúng

tôi thực hiện nghiên cứu chế tạo mô hình mô phỏng súng máy tự động đơn giản sử dụng công nghệ in 3D và để xuất một thuật toán xử lý ảnh giúp phát hiện, nhận dạng và tự động bám sát mục tiêu dựa trên YOLOv8. Phần tiếp theo của bài báo, chúng tôi sẽ giới thiệu tổng quan về giải pháp bao gồm sơ đồ khối hệ thống. Phần 3 giới thiệu tóm tắt về thiết kế cơ khí trước khi thực hiện in in 3D mô hình từ nhựa tái chế PLA, phần điện (động cơ, cảm biến), phần xử lý ảnh (máy tính, camera) cũng như thuật toán điều khiển cùng phương pháp xây dựng tập mẫu. Phần cuối cùng đưa ra những đánh giá tổng kết về những giá trị đạt được sau nghiên cứu, đồng thời đưa ra phương hướng phát triển tiếp theo trong tương lai.

2. TỔNG QUAN VỀ GIẢI PHÁP CÔNG NGHỆ

2.1. Sơ đồ khối hệ thống

Sơ đồ khối tổng quan của hệ thống (hình 1) gồm hai khối chính: Khối máy tính điều khiển và khối mô hình mô phỏng súng máy.

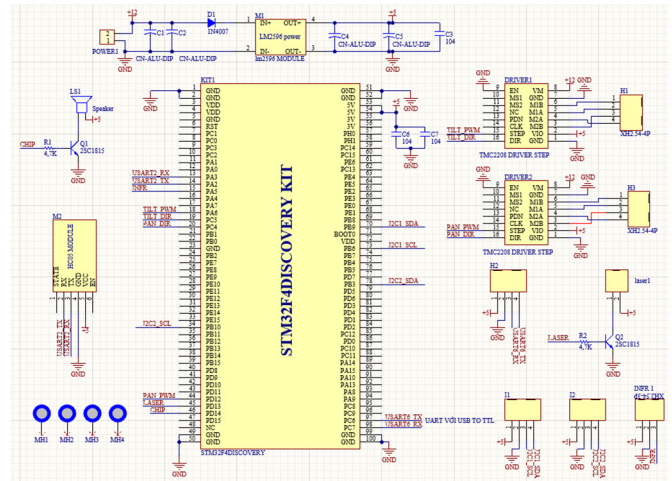


Hình 1. Sơ đồ tổng quan hai khối chính của hệ thống

Khối máy tính điều khiển: Là máy tính cá nhân, máy tính nhúng hoặc máy tính công nghiệp có nhiệm vụ điều khiển mô hình súng máy theo những chế độ điều khiển khác nhau bằng phần mềm điều khiển và giám sát được viết bằng ngôn ngữ Python. Máy tính sẽ nhận hình ảnh được gửi về từ camera trên mô hình mô phỏng súng máy thông qua đường truyền dữ liệu USB, thực hiện xử lý dữ liệu hình ảnh với mô hình YOLOv8. Sau khi đã xác định được tọa độ của vật thể trong không gian, máy tính sẽ gửi tọa độ đến mạch điều khiển mô hình mô phỏng súng máy thông qua đường truyền UART để thực hiện quá trình bắt bám đối tượng.

Khối mô hình mô phỏng súng máy: Là mô hình mô phỏng hệ thống súng máy trong thực tế được, gồm mô hình súng máy được in 3D, camera và mạch điều khiển. Mô hình hoạt động bắt bám đối tượng dựa trên tọa độ của đối tượng nhận được từ khối máy tính điều khiển bởi mạch điều khiển. Mạch điều khiển mô hình,

với sơ đồ mạch nguyên lý như ở hình 2, sử dụng vi điều khiển chính là STM32F411VET6 kết hợp với cảm biến gia tốc GY-521 6 trục IMU MPU6050 nhằm xác định các góc xoay theo phương ngang (Pan) và góc xoay theo phương nghiêng (Tilt), cảm biến hồng ngoại TCRT5000 xác định vị trí ban đầu của súng khi khởi động, module điều khiển động cơ bước TMC2208 điều khiển hai động cơ bước và một màn hình LCD2004 để hiển thị các chế độ hoạt động, thời gian hoạt động, các góc quay giúp người dùng có cái nhìn trực quan.

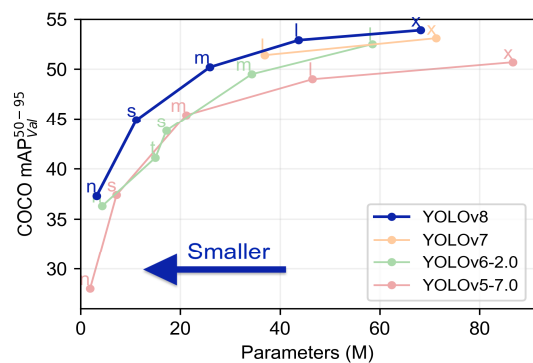


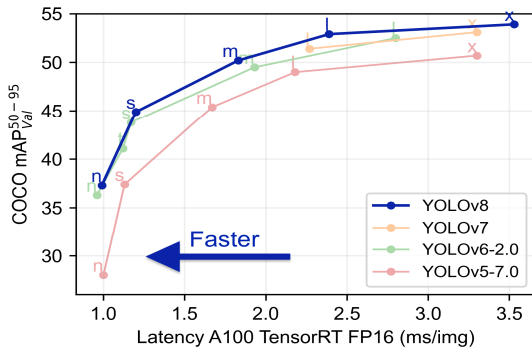
Hình 2. Sơ đồ nguyên lý mạch điều khiển mô hình

2.2. Mô hình mạng YOLOv8

YOLO [11] là một mô hình mạng nơ-ron tích chập được ứng dụng nhiều cho việc nhận dạng đối tượng và phân loại thời gian thực bởi mô hình này có ưu điểm là có thể duy trì được độ chính xác ở mức ổn định mặc dù tốc độ của nó so với các cấu trúc khác là nhanh hơn rất nhiều.

Có rất nhiều thuật toán YOLO được phát triển và công bố từ năm 2015 đến nay, trong đó mô hình mới nhất do Ultralytics phát triển là YOLOv8[12]. Mô hình YOLOv8 được phát triển dựa trên mô hình YOLO phiên bản trước với nhiều cải tiến làm cho mô hình trở lên nhanh hơn, chính xác hơn (hình 3) và dễ dàng sử dụng hơn khi được đóng gói dưới dạng thư viện trong Python.





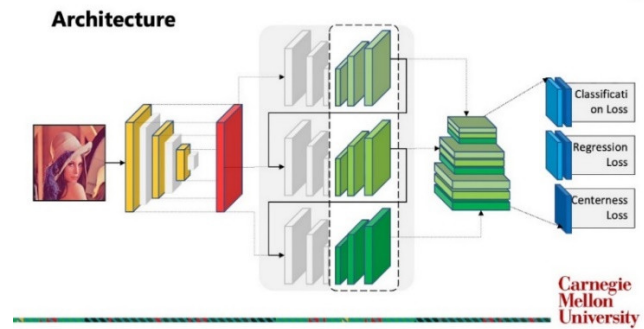
Hình 3. So sánh tham số và tốc độ của mô hình YOLOv8 so với các phiên bản trước [13]

Mô hình YOLOv8 được hình thành dựa trên hai phần chính là "Backbone" và "Head". Backbone được thiết kế dựa trên kiến trúc mạng CNN với Darknet-53. "Backbone" trích xuất những đặc trưng ảnh đầu vào thông qua 5 giai đoạn, mỗi giai đoạn sẽ bao gồm một số khối "Bottleneck" được kết nối với nhau. Trong đó, khối "Bottleneck" được xây dựng bởi một khối 3x3 và một khối 1x1 "convolutional layer". Từ đó, mỗi một giai đoạn trong "Backbone" sẽ trích xuất ra những đặc trưng ở độ phân giải khác nhau nhờ việc giảm kích thước các đầu vào thông qua các khối C2F (Convolution 2x2 follower by fusion). Các đặc trưng được trích xuất ra từ Backbone sẽ được sử dụng trong phần "Head" để dự đoán vị trí, kích thước và lớp của đối tượng trong ảnh đầu vào thông qua việc sử dụng ba nhánh song song để nhận diện được đối tượng ở những độ phân giải khác nhau.

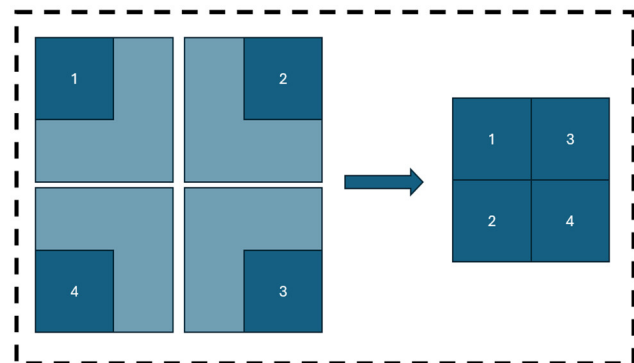
Trong khi các mô hình cũ sử dụng những "neo" (Anchor box) để dự đoán hộp giới hạn (Bounding Box) cho vật thể thì mô hình YOLOv8 sử dụng Anchor Free Detection - phương pháp phát hiện đối tượng trong ảnh để xác định trực tiếp hộp giới hạn cho vật thể mà không cần sử dụng các "neo". Đầu tiên, phương pháp này sẽ bắt đầu đưa hình ảnh đầu vào qua một mạng CNN để trích xuất từ hình ảnh các đặc trưng ở các tỷ lệ không gian khác nhau. Sau đó, áp dụng kiến trúc mạng neural FPN để xây dựng những mô hình đa cấp (Multi-scale feature representations) nhằm mục đích tạo ra một tập hợp các đặc trưng với độ phân giải khác nhau từ một ảnh đầu vào duy nhất. Tiếp đến, "Anchor Free Detection" với cấu trúc như ở Hình 4 sẽ dự đoán trực tiếp các Bounding Box với tọa độ tâm (x, y), kích thước (w, h) và xác suất đối tượng trên điểm lưới được dự đoán và tính toán dựa trên độ lệch của mỗi ô lưới trong "Feature map" so với vị trí của ô lưới đó.

Trong mô hình YOLOv8 cũng ứng dụng một kỹ thuật để tăng cường dữ liệu là "Mosaic Data Augmentation" nhằm mục đích cải thiện hiệu suất và độ chính xác của mô hình bằng cách tăng cường tính đa dạng của dữ liệu huấn

luyện. Kỹ thuật này, như mô tả ở hình 5, sẽ bắt đầu bằng việc chọn 4 ảnh ngẫu nhiên từ tập dữ liệu rồi sau đó mỗi ảnh sẽ được cắt thành 4 phần nhỏ với kích thước bằng nhau rồi từ đó lấy mỗi một phần nhỏ từ mỗi ảnh để sắp xếp ngẫu nhiên ra một ảnh mới. Điều này giúp tăng cường tính đa dạng của dữ liệu huấn luyện bên cạnh đó cũng tránh xảy ra hiện tượng "Overfitting" trên mô hình khiến cho khả năng khái quát hóa của mô hình được tăng cao.



Hình 4. Cấu trúc phương pháp phát hiện đối tượng Anchor Free Detection [14]



Hình 5. Kỹ thuật Mosaic Data Augmentation

3. NHẬN DẠNG VÀ ĐIỀU KHIỂN BẨM ĐỐI TƯỢNG

3.1. Huấn luyện mô hình YOLOv8 trên tập dữ liệu cá nhân

Quá trình đầu tiên trong xử lý ảnh và huấn luyện là thu thập dữ liệu về đối tượng và tiền xử lý các dữ liệu về đối tượng. Dữ liệu thu thập càng đa dạng, chính xác thì mô hình đầu ra có hiệu suất càng cao. Trong bài báo này, chúng tôi sử dụng đối tượng để nhận dạng minh họa như ở hình 6.

Trong quá trình thu thập dữ liệu từ thực tế, chúng tôi đã tổng hợp được hơn 1000 ảnh. Trong đó, như ở bảng 1, bộ dữ liệu được chia ra thành hai tập dữ liệu "train set" và "valid set" tương ứng với 80% và 20% dữ liệu gốc. Trong quá trình tiền xử lý dữ liệu, chúng tôi sử dụng công cụ gán nhãn đối tượng trên trang Roboflow.com và thực hiện huấn luyện mô hình sử dụng Google Colab. Cuối cùng, đánh giá lại hiệu suất của mô hình thông qua thông số mAP.

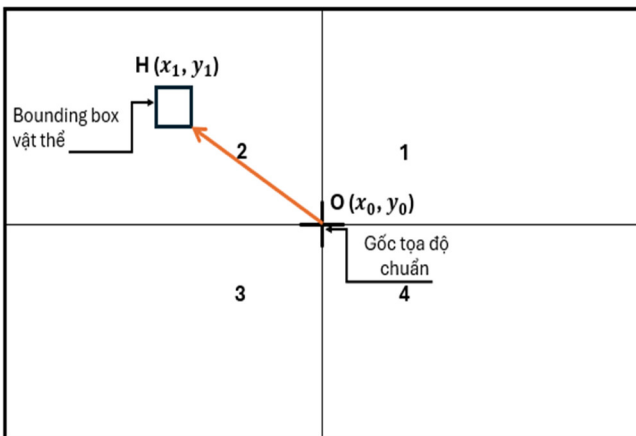


Hình 6. Đối tượng được nhận dạng

Bảng 1. Bảng chia tập dữ liệu cá nhân (Dataset Split)

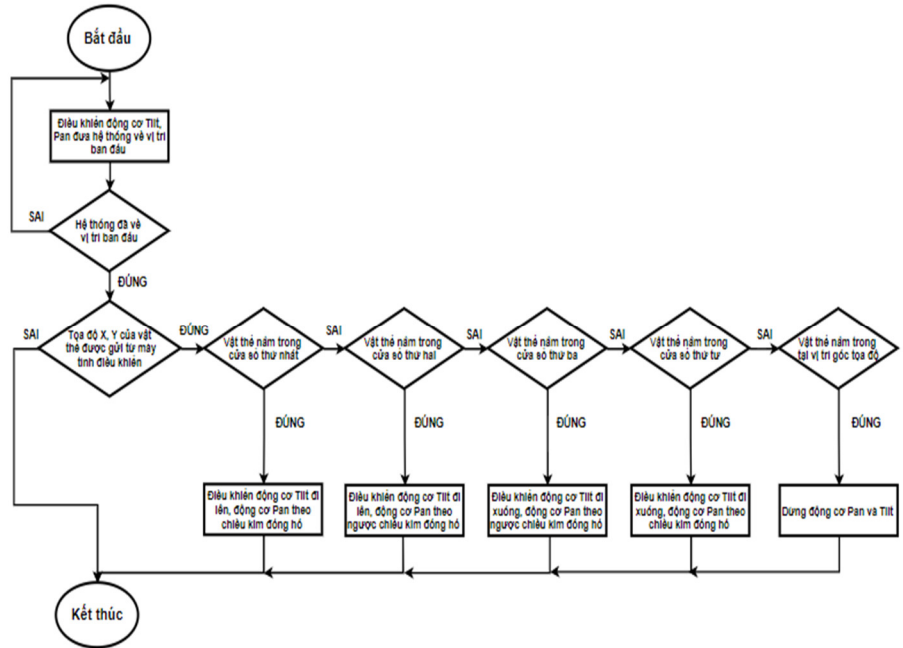
| Loại dữ liệu | Số lượng ảnh |
|--------------|--------------|
| Train Set | 896 |
| Valid Set | 224 |
| Tổng | 1120 |

3.2. Phương pháp điều khiển chuyển động



Hình 7. Góc tọa độ chuẩn và tọa độ của vật thể

Mạch điều khiển sẽ điều khiển mô hình súng máy tự động bám theo vật thể dựa vào tọa độ của vật thể do máy tính điều khiển gửi tới. Tọa độ này được quy ước theo vị trí điểm ảnh (Pixel) trên màn hình (hình 7). Màn hình được phân làm 4 góc phần tư nhỏ hơn được quy ước có số thứ tự tăng dần theo chiều ngược chiều kim đồng hồ với vị trí tâm của màn hình cũng là gốc tọa độ chuẩn $O(x_0, y_0)$, $x_0 = 0$ và $y_0 = 0$. Nếu vị trí $H(x_1, y_1)$ của vật thể đang được phát hiện nằm trong góc phần tư thứ nhất thì x_1 sẽ mang giá trị dương và y_1 sẽ mang giá trị dương, góc phần tư thứ hai x_1 sẽ mang giá trị âm và y_1 sẽ mang giá trị dương, góc phần tư thứ ba x_1 sẽ mang giá trị âm và y_1 sẽ mang giá trị âm và góc phần tư thứ tư x_1 sẽ mang giá trị dương và y_1 sẽ mang giá trị âm. Dựa trên đó, mô hình sẽ xác định được rõ hướng di chuyển để có thể bám theo đối tượng sao cho tâm vật thể tiệm cận tới gốc O.

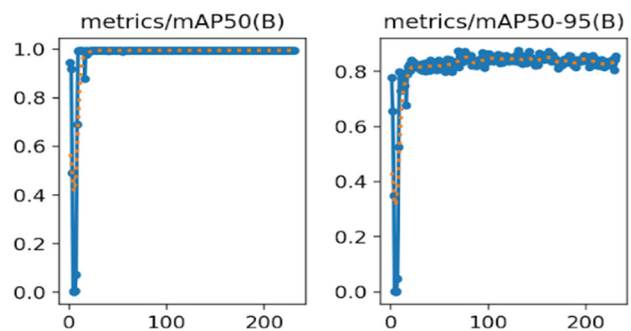


Hình 8. Lưu đồ thuật toán điều khiển mô hình tám theo vật thể

Với quy ước động cơ bước điều khiển mô hình súng theo phương ngang là động cơ Pan, động cơ bước điều khiển nòng súng của mô hình theo phương nghiêng là động cơ Tilt, thì lưu đồ thuật toán điều khiển mô hình tám theo vật thể được mô tả như ở hình 8. Sau khi khởi động, hệ thống sẽ tự động di chuyển về vị trí ban đầu (vị trí sẵn sàng để vào các chế độ hoạt động). Khi đã về được vị trí ban đầu, mạch điều khiển sẽ liên tục kiểm tra tọa độ của vật thể được gửi từ máy tính điều khiển đến và thực hiện điều khiển các động cơ Pan và Tilt di chuyển hướng súng đến vị trí của vật thể (vị trí $H(x_1, y_1)$) trong không gian. Khi gốc tọa độ chuẩn $O(x_0, y_0)$ trùng với $H(x_1, y_1)$ thì sẽ thực hiện dừng các động cơ Pan và Tilt.

4. ĐÁNH GIÁ KẾT QUẢ

4.1. Đánh giá mô hình YOLOv8 trên tập dữ liệu cá nhân



Hình 9. Biểu đồ mAP50 và mAP50-95

Sau khi huấn luyện mô hình trên Google Colab với 231 epochs, cho thấy kết quả tốt nhất với $mAP50 = 0,997$ và $mAP50-95 = 0,867$ (hình 9). Trong đó, mAP50 là độ chính xác trung bình (mean Average Precision) được tính toán

ở ngưỡng IoU = 0,5 thể hiện độ chính xác của mô hình khi phát hiện các đối tượng có kích thước lớn và dễ nhìn thấy, mAP50-95 là độ chính xác trung bình ở các mức IoU khác nhau thể hiện độ chính xác của mô hình khi phát hiện các đối tượng ở các kích thước khác nhau.

Mô hình đạt được độ chính xác cao ở cả mAP50 và mAP50-95 cho thấy rằng mô hình đạt được độ chính xác cao khi phát hiện các đối tượng có kích thước lớn cũng như là các đối tượng có nhiều kích thước khác nhau khi ở gần hoặc ở xa.

4.2. Đánh giá kết quả mô hình bắt bám đối tượng



Hình 10. Mô hình mô phỏng súng máy và máy tính điều khiển

Hình 10 mô tả mô hình mô phỏng súng máy và máy tính điều khiển điều khiển sau hoàn thiện, trong đó các khối tương ứng

- (1): Khối Camera
- (2): Động cơ bước điều khiển phương nghiêng (Tilt)
- (3): Động cơ bước điều khiển phương ngang (Pan)
- (4): Cảm biến MPU6050 và hộp bảo vệ
- (5): Màn hình LCD hiển thị thông số trực tiếp của thiết bị
- (6): Hộp mạch điều khiển mô hình
- (7): Máy tính điều khiển.

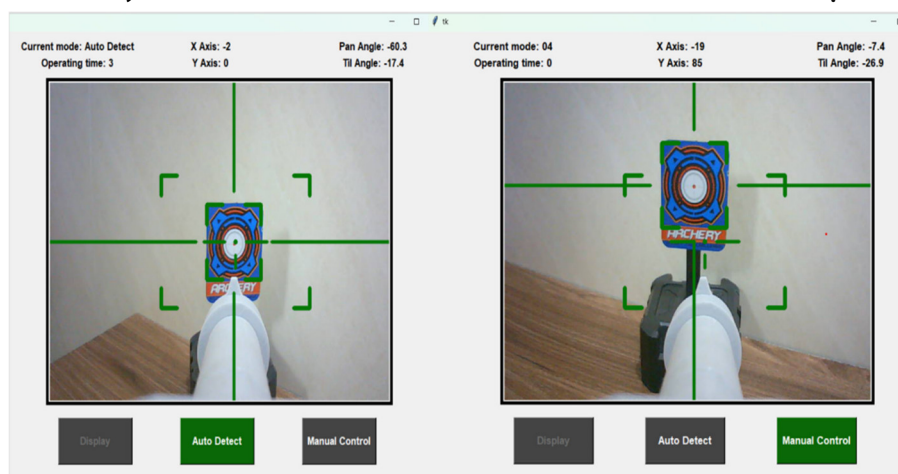
Camera nằm phía trên cùng súng giúp cho hệ có được tầm nhìn xa và rộng nhất có thể. Vị trí nằm ở phía trên cùng sẽ làm cho hệ thống có thể bao quát một cách toàn cảnh với một khoảng không gian rộng lớn mà không bị che khuất tầm nhìn bởi các bộ phận khác. Đồng thời đầu ngắm của súng cũng được thiết kế để có thể xuất hiện trong khung hình, nguyên lý này giống với ống ngắm trên các loại súng trường, súng bắn tỉa giúp cho việc ghim nòng súng vào tâm của mục tiêu một cách dễ dàng hơn. Tuy nhiên, trong bài báo này, chúng tôi chỉ tập trung vào để xuất các thuật toán điều khiển bám đối tượng còn việc nghiên cứu đường bay quỹ đạo của đạn hay động học súng máy chưa được đề cập đến.

Hình 11 mô tả màn hình điều khiển giám sát mô hình súng máy bao gồm hai chế độ tương ứng là chế độ tự động bám theo vật thể (Auto Detect) và chế độ điều khiển bằng tay thông qua các phím trên bàn phím máy tính (Manual Control), khi người sử dụng lựa chọn chế độ nào, nút mang tên tương ứng bên dưới sẽ chuyển sang màu xanh.

Tại chế độ tự động bám theo vật thể (màn hình hiển thị bên trái), hệ thống sẽ tự động nhận diện và điều chỉnh vị trí sao cho nòng súng luôn hướng thẳng vào vật thể, chiếu theo gốc tọa độ thì vật thể luôn nằm trên gốc tọa độ O (x_0, y_0). Kết quả cho thấy hệ thống đã tự động bắt, bám đối tượng tốt với thời gian đáp ứng nhanh khi đối tượng di chuyển với tốc độ trong khoảng 0.3-0.5 m/s. Như hiển thị ở giao diện hình 11, hệ thống bắt bám đối tượng tới tọa độ thực trên hệ trục tọa độ Oxy là H (-2, 0).

Tại chế độ điều khiển bằng tay thông qua 4 phím A,W,D,S trên bàn phím máy tính (màn hình hiển thị bên phải), tọa độ thực của vật thể đang hiển thị trên hệ trục tọa độ Oxy là H (-19, 85) nằm ở góc phần tư thứ 2.

5. KẾT LUẬN



Hình 11. Màn hình giám sát và điều khiển mô hình súng máy

Bài báo đề xuất sử dụng YOLOv8 nhằm phát hiện và tự động bám theo mục tiêu cùng một số chức năng khác như điều khiển qua nút điều hướng A,W, D, S hay tự động về vị trí ban đầu khi hệ được cấp nguồn. Độ chính xác nhận dạng mục tiêu là 97,86%, cho thấy thuật toán đề xuất trong ứng dụng công nghệ xử lý ảnh vào trong mô hình súng máy phát hiện và bám sát mục tiêu tự động ứng dụng cho máy bay quân sự thực tế có tính khả thi. Trong thời gian tới, chúng tôi sẽ

cải thiện thuật toán để có thể nhận dạng trong trường hợp môi trường có lẫn các đối tượng khác tương đồng, tối ưu hơn trong việc bám và khóa mục tiêu cho trước, cũng như thêm chức năng điều khiển súng theo điều hướng bằng mũ của người trên máy bay.

TÀI LIỆU THAM KHẢO

- [1]. Vig S., Arora A., Arya G., "Automated license plate detection and recognition using deep learning," In *International Conference on Advancements in Interdisciplinary Research*, Cham: Springer Nature Switzerland, 419-431, 2022.
- [2]. Munawar H.S., Khalid U., Maqsood A., "Fire detection through Image Processing; A brief overview." In *Int. Conf. Cult. Technol*, 203-208, 2017.
- [3]. Mahanthesha U., "Human Motion Detection and Tracking for Real-Time Security System," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCCE)*, 2016.
- [4]. Iqbal J., Pasha S.M., Baizid K., Khan A.A., Iqbal J., "Computer vision inspired real-time autonomous moving target detection, tracking and locking," *Life Science Journal*, 10(4), 3338-3345, 2013.
- [5]. Iqbal J., Pasha M., Riaz-un-Nab, H.K., Iqbal J., "Real-time target detection and tracking: A comparative in-depth review of strategies" *Life Science Journal*, 10(3), 804-813, 2013.
- [6]. Shujaat Khan, Huan Liu, Ralph Chatham, Charles Jacobus, David Caulfield, "Real-Time Autonomous Target Tracking and Engagement with a Sniper Rifle System," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2008.
- [7]. D. A. Bell, D. J. Coleman, S. K. Khan, A. L. Stokes, "Target Tracking and Engagement System," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2010.
- [8]. Defense Express, Ukraine Gets Unique Slinger Anti-Drone Weapon Station From Australian EOS, 2023. https://en.defence-ua.com/weapon_and_tech/ukraine_gets_unique_slinger_anti_drone_weapon_station_from_australian_eos-7818.html
- [9]. <https://baonghean.vn/viet-nam-san-xuat-sung-may-dieu-khien-tu-xa-post152984.html>
- [10]. <https://www.qdnd.vn/cong-nghiep-quoc-phong-4-0/nghien-cuu-va-san-xuat-khi-tai-quang-dien-tu-tai-viet-nam-683910>

[11]. J. Redmon, A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[12]. Rizwan Munawar, Glenn-Jocher, AyushExel, *YOLO: A Brief History*, <https://docs.ultralytics.com/>

[13]. <https://github.com/ultralytics/ultralytics>

[14]. <https://github.com/shayeree96/11785-CMU---FCOS-with-MLFPN>

AUTHORS INFORMATION

Dinh Tuan Anh, Nguyen Tien Hoang, Ha Viet Anh, Vu Huu Thich, Pham Van Hung

Hanoi University of Industry, Vietnam