

Ứng dụng mô hình học sâu trong thị giác máy tính cho hệ bắt-bám mục tiêu của khí cụ bay tự dẫn vắc vai huấn luyện

Đặng Võ Công*, Lê Thị Hằng, Hoàng Huy Lê, Phạm Tuấn Hùng

Viện Kỹ thuật Phòng không - Không quân, số 166 Hoàng Văn Thái, Thanh Xuân, Hà Nội, Việt Nam.

*Email: mrcong.vkt@gmail.com

Nhận bài: 27/01/2024; Hoàn thiện: 11/3/2024; Chấp nhận đăng: 08/4/2024; Xuất bản: 22/4/2024.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.94.2024.159-165>

TÓM TẮT

Bài báo trình bày nội dung nghiên cứu ứng dụng mô hình học sâu trong thị giác máy tính cho hệ bắt-bám mục tiêu của khí cụ bay tự dẫn vắc vai huấn luyện. Mục đích của nghiên cứu là sử dụng hệ camera số như một hệ bắt-bám giả lập có tính năng phù hợp với tính năng của đầu tự dẫn vắc vai, làm cơ sở xây dựng các khí cụ bay tự dẫn vắc vai huấn luyện. Để bắt-bám được mục tiêu di động ngoài thực địa, khả năng nhận dạng mục tiêu 3D phải đáp ứng yêu cầu về độ chính xác, tuy nhiên, khi áp dụng thuật toán xử lý ảnh truyền thống đã không hiệu quả. Để giải quyết vấn đề này, bài báo tập trung nghiên cứu ứng dụng mô hình học sâu YOLO cho hệ bắt-bám camera số. Trên cơ sở sử dụng mô hình YOLO phiên bản 8, nhóm nghiên cứu đã thử nghiệm thu thập dữ liệu, xử lý hình ảnh, huấn luyện mô hình và xem xét khả năng bắt-bám mục tiêu của một hệ camera số. Đồng thời, chế thử thiết bị khí cụ bay tự dẫn vắc vai huấn luyện cỡ 72 mm với hệ bắt-bám camera số và thử nghiệm ngoài thực địa. Kết quả, khả năng bắt-bám mục tiêu của hệ camera số phù hợp theo chức năng đầu tự dẫn hồng ngoại được xây dựng trên cơ sở con quay điện tử trong phạm vi giới hạn về khoảng cách.

Từ khóa: Hệ bắt-bám mục tiêu; Hệ camera số; Mô hình học sâu; Thuật toán YOLO.

1. MỞ ĐẦU

Ngày nay, công nghệ thị giác máy tính và học sâu tiếp tục đóng một vai trò quan trọng để giải quyết các bài toán về nhận dạng, phát hiện và theo dõi đối tượng, định vị đối tượng trong hình ảnh và video, được quan tâm trong nhiều lĩnh vực khác nhau như hệ thống giám sát an ninh, môi trường, xe tự hành, rô bốt, và chăm sóc sức khỏe [1]. Thực tế, thị giác máy tính cũng đã được ứng dụng hiệu quả để giải quyết nhiều bài toán, như trong thể thao, nông nghiệp, giao thông [2-4], phương tiện bay không người lái (UAV) [5].

Bài báo này tập trung nghiên cứu ứng dụng mô hình học sâu trong thị giác máy tính cho hệ bắt-bám mục tiêu của khí cụ bay tự dẫn vắc vai huấn luyện (KCBVV-HL), trong mối quan tâm về ứng dụng thị giác máy tính nhằm giảm chi phí sản xuất và thay thế hệ bắt-bám hồng ngoại bởi hệ bắt-bám camera số. Nội dung nghiên cứu xuất phát từ nhu cầu thực tiễn, sự đòi hỏi về trang bị học tập trong huấn luyện khí cụ bay tự dẫn vắc vai (KCBVV). Thực tế, các hệ bắt-bám hồng ngoại trên cơ sở con quay quang điện tử hiện đang được sử dụng trên các KCBVV-HL là nguồn vật tư khan hiếm và có giá thành cao. Giải quyết vấn đề này, camera số là sản phẩm khá phổ biến, chi phí rẻ, có thể kết hợp với máy tính cả phần cứng và phần mềm để tạo ra hệ bắt-bám mục tiêu camera số ứng dụng cho KCBVV-HL giả lập theo chức năng bắt-bám mục tiêu đầu tự dẫn hồng ngoại của KCBVV trong giai đoạn ngắm bắn.

Xem xét đối tượng cần được bắt-bám là như một vật thể 3D bay ngoài trời, nhóm nghiên cứu đã lựa chọn mô hình học sâu YOLO [6], bởi thế mạnh về tốc độ suy luận, cho phép phát hiện đối tượng theo thời gian thực, kiến trúc mạng đơn giản và bộ dữ liệu huấn luyện tối giản [7]. Thực tế, áp dụng thuật toán Lọc tương quan Kernel hóa (Kernelized Correlation Filter) [8], đã cho thấy không hiệu quả trong bắt-bám mục tiêu 3D khi ở khoảng cách trên 10 mét, khung nền ánh sáng thay đổi theo điều kiện thời tiết, mô hình mục tiêu nghiêng ngã và thay đổi hình dạng mẫu ngẫu nhiên theo các điều kiện môi trường khác nhau.

2. CƠ SỞ THIẾT LẬP HỆ BẮT-BÁM VÀ MÔ HÌNH HỌC SÂU THỊ GIÁC MÁY TÍNH

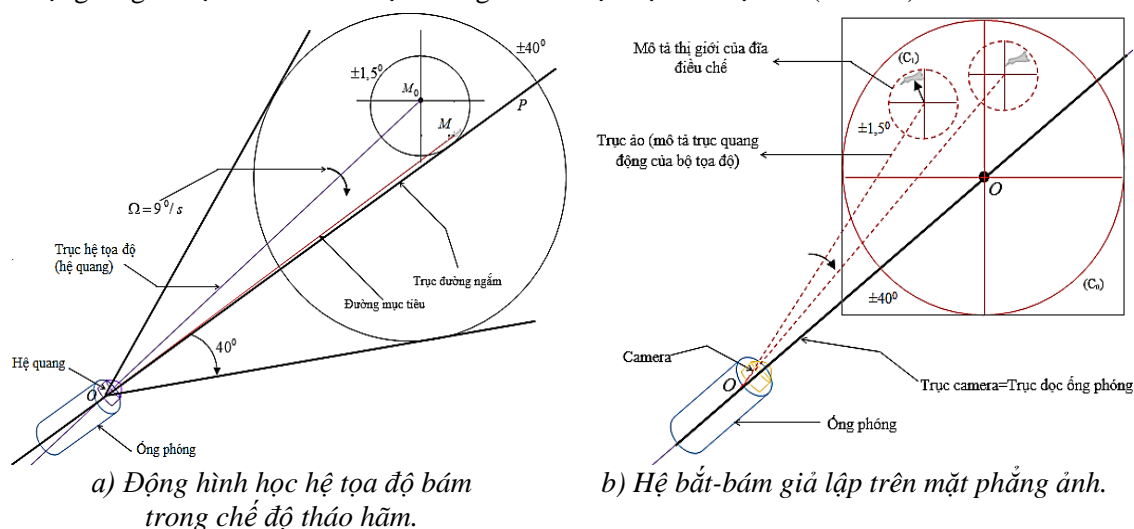
2.1. Cơ sở xây dựng hệ bắt-bám camera số

Trong nghiên cứu này, hệ bắt-bám camera số được phỏng theo KCBVV-HL cỡ 72 mm [9].

Theo nguyên lý dẫn tỷ lệ, về bản chất các đầu tự dẫn quang thụ động và các đầu tự dẫn hồng ngoại là các hệ tọa độ bám [10]. Để thực hiện bắt và bám mục tiêu, hệ tọa độ bám phải luôn bắt và bám mục tiêu, duy trì sao cho trục quang hệ tọa độ bám luôn trùng với đường ngắm khí cụ bay - mục tiêu (hay hướng của véc tơ cự li). Trong đó, hoạt động của hệ tọa độ bám gồm hai chế độ làm việc chính:

Chế độ hãm: Hệ tọa độ bám chưa được giải phóng. Trục của hệ tọa độ bám được duy trì trùng với trục đầu tự dẫn, khi đó trường nhìn của đầu tự dẫn bị giới hạn và là thị giới của đĩa điều chế. Đối với KCBVV-HL cỡ 72 mm, thị giới của đĩa điều chế là $\pm 1,5^\circ$, trong đó đĩa điều chế được gắn trên hệ tọa độ bám.

Chế độ tháo hãm: Khi đủ các điều kiện cho phép đầu tự dẫn bắt được mục tiêu, trặc thủ bóp cò nắc 1 thì hệ tọa độ bám được tháo hãm và chuyển sang chế độ bám sát, cho phép mở rộng trường nhìn của đầu tự dẫn. Do thị giới của đĩa điều chế được gắn trên hệ tọa độ bám có thể tịnh tiến bám theo mục tiêu theo thông tin về góc lệch và hướng. Đối với KCBVV-HL cỡ 72 mm trường nhìn mở rộng có giới hạn $\pm 40^\circ$ theo mọi hướng so với trục dọc đầu tự dẫn (hình 1a).



Hình 1. Thiết lập hệ bắt-bám giả lập trên hệ camera số phỏng theo hệ tọa độ bám.

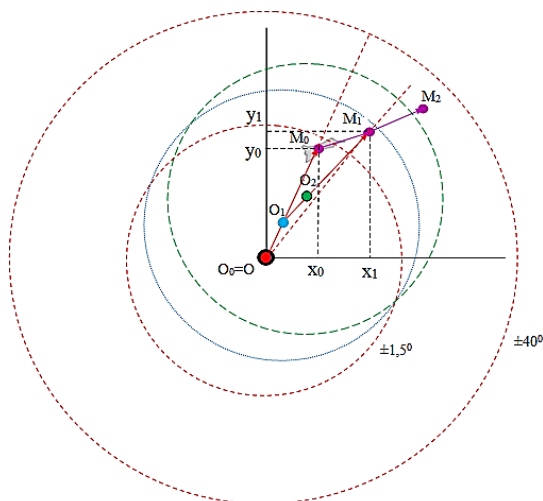
Trên cơ sở đó, hệ bắt-bám camera số được phỏng theo các đặc điểm của hệ tọa độ bám phù hợp với tính năng bắt và bám mục tiêu của đầu tự dẫn trên ống phóng (hình 1b). Hệ bắt-bám camera số cần tạo giả được trường nhìn với vòng tròn thị giới ảo $\pm 1,5^\circ$ mô tả thị giới của đĩa điều chế ở chế độ hãm và chuyển động trong trường nhìn mở rộng $\pm 40^\circ$ ở chế độ tháo hãm.

Theo hình 1b, trong mặt phẳng ảnh của camera tạo một trường nhìn ảo $\pm 1,5^\circ$ (mô tả như đường tròn C_1 có tâm trùng với tâm O tại thời điểm tháo hãm). Đường tròn này sẽ mô phỏng thị giới của đĩa điều chế hệ tọa độ bám chuyển động tịnh tiến trong trường nhìn mở rộng $\pm 40^\circ$.

Về nguyên lý, hệ tọa độ bám gồm vật kính, đĩa điều chế và quang trở là thành phần thu tín hiệu hồng ngoại phát ra từ mục tiêu. Tín hiệu hồng ngoại được chuyển đổi thành các chùm xung mang thông tin về độ lệch và hướng mục tiêu. Tín hiệu này, được đưa đến hệ thống hiệu chỉnh con quay để tạo lệnh làm trùng trục quang bộ tọa độ với đường ngắm khí cụ bay - mục tiêu theo chuyển động tịnh tiến với vận tốc đến 9 [°/s] [10]. Để phù hợp cho hệ bắt-bám camera số, giả thiết tâm nhiệt trùng với tâm hình học của mục tiêu.

Về mặt ý tưởng thị giác máy tính, sai lệch và hướng mục tiêu được quy về tọa độ ảnh, sau nhận dạng mục tiêu, thuật toán sẽ xác định vị trí và sai lệch tọa độ tâm mục tiêu và tâm (C_1). Theo sai lệch, đường tròn (C_1) sẽ được lập trình chuyển động tịnh tiến về phía tâm mục tiêu với tốc độ 9 $[^{\circ}/s]$, trên cơ sở áp đặt thời gian dịch chuyển (t_{dc}) thay đổi theo sai lệch. Bên cạnh đó, để mô phỏng chuyển động của đĩa điều chế (trường nhìn $\pm 1,5^{\circ}$) trong mặt phẳng vuông góc với trục dọc của đầu tự dẫn, cần tạo một vòng tròn thị giới ảo trên mặt phẳng ảnh của camera có tính chất chuyển động giống như chuyển động của đĩa điều chế trên hệ tọa độ đầu tự dẫn (hình 2).

Theo hình 2, tại thời điểm chưa tháo hãm, vòng tròn thị giới ảo ($\pm 1,5^{\circ}$) có tâm O_0 trùng với tâm O . Khi hệ camera bắt được mục tiêu và bắt đầu tháo hãm, vòng tròn thị giới ảo chuyển động theo đường thẳng về phía ảnh mục tiêu M_0 với tốc độ dịch chuyển điểm ảnh N pixels/s (trương đương với 9 $[^{\circ}/s]$ là tốc độ tiến động của con quay trên đầu tự dẫn). Giá trị N phụ thuộc vào tiêu cự ống kính và kích thước điểm ảnh của cảm biến ảnh. Cứ như vậy sau những khoảng thời gian khá nhỏ bằng thời gian lấy mẫu của thuật toán xử lý ảnh, vòng tròn thị giới ảo chuyển động về phía ảnh mục tiêu M_1, M_2, \dots trong trường nhìn có giới hạn $\pm 40^{\circ}$. Tại các thời điểm xử lý ảnh, hoàn toàn có thể tính toán được tọa độ tâm và tọa độ của vòng tròn thị giới ảo trong mặt phẳng ảnh. Trong nghiên cứu này, 1° trương đương 74 Pixels.



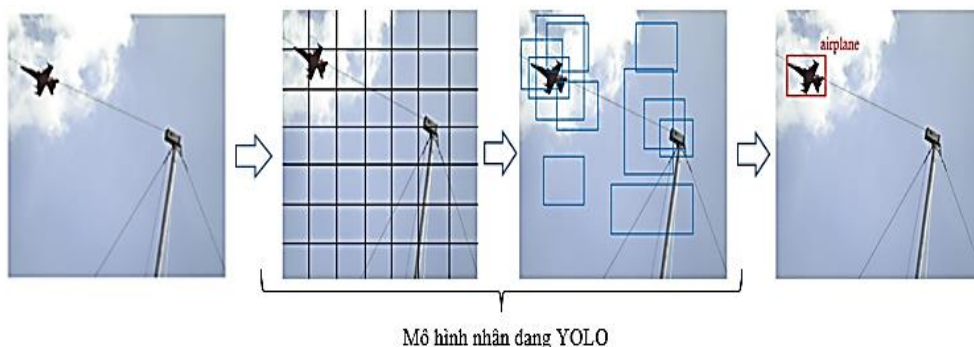
Hình 2. Thể hiện chuyển động tịnh tiến của vòng tròn thị giới ảo trên mặt phẳng ảnh.

Qua phân tích hoạt động của hệ tọa độ bám đầu tự dẫn KCBVV cỡ 72 mm, nhóm nghiên cứu sử dụng mô hình học sâu YOLO trong thị giác máy tính để giải quyết bài toán bắt-bám mục tiêu, được trình bày ở phần dưới đây.

2.2. Mô hình YOLO và cách huấn luyện

YOLO là một mô hình mạng nơ-ron tích chập CNN cho phép phát hiện, nhận dạng và phân loại đối tượng. YOLO được tạo ra từ việc kết hợp giữa các lớp tích chập để trích xuất các đặc trưng ảnh và các lớp kết nối để dự đoán xác suất và tọa độ của đối tượng.

Đầu vào của mô hình là một bức ảnh, được chia thành nhiều ô có kích thước $(2n+1) \times (2n+1)$, với $n = 1, 2, 3, \dots$ Mỗi ô sẽ chứa thông tin về tọa độ và kích thước của hộp giới hạn chứa đối tượng (hay mục tiêu cần bắt-bám), khả năng và xác suất xuất hiện mục tiêu. Mô hình sẽ dự đoán và đưa ra quyết định là có hay không mục tiêu cần nhận dạng trong khung hình và xác định tọa độ của mục tiêu.

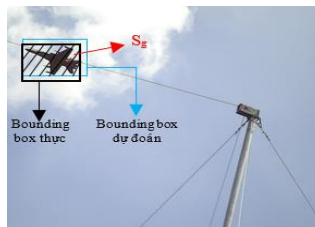


Hình 3. Hoạt động của mô hình học sâu YOLO.

Để tính được độ chính xác của hộp giới hạn dự đoán cần sử dụng phép đo IoU (Intersection over Union). Chỉ số IoU là tỷ lệ đo lường mức độ giao nhau giữa 2 hộp giới hạn thực và hộp giới hạn dự đoán, được xác định:

$$IoU = \frac{S_g}{S_h} \quad (1)$$

Trong đó: S_g - Diện tích phần giao nhau giữa hộp giới hạn thực và hộp giới hạn dự đoán; S_h - Diện tích phần hợp của hộp giới hạn thực và hộp giới hạn dự đoán.

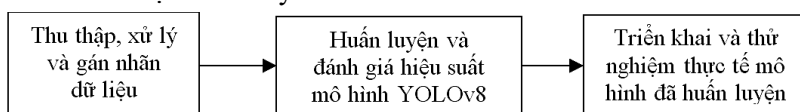


Hình 4. Mô tả diện tích giao nhau S_g .

Về lý thuyết, mỗi sự nhận dạng về mục tiêu được đánh giá thông qua một giá trị IoU riêng từ 0 đến 1. Khi $IoU > 0,5$ mục tiêu có thể xem xét là đã được nhận dạng chính xác [7]. Song thông qua thực nghiệm, để đảm bảo chất lượng bắt-bám mục tiêu, nhóm nghiên cứu đã chọn ngưỡng chỉ tiêu nhận dạng là $IoU_{th} \geq 0,7$ cho bài toán.

YOLO dự đoán nhiều hộp giới hạn cho mỗi ô lưới, để tính lỗi cho các dự đoán đúng, một hộp giới hạn có IoU cao nhất chứa đối tượng thực được xác định. Từ hộp giới hạn xác định, trong mô hình YOLO sử dụng hàm Sum-Squared Error (SSE) để tính mất mát [7]. Hàm mất mát được tự động tính toán sau mỗi vòng huấn luyện, quá trình huấn luyện sẽ kết thúc khi hàm mất mát đạt giá trị đạt cực tiểu.

Trình tự các bước của thuật toán YOLO nhận dạng mục tiêu cho hệ bắt-bám camera số của KCBVV-HL cỡ 72 mm được trình bày như hình 5.



Hình 5. Trình tự các bước của thuật toán YOLO.

Bước 1: Thu thập, xử lý và gán nhãn dữ liệu

Thu thập các video mục tiêu từ hệ camera số ngoài thực địa, cắt thành các khung hình để tạo bộ dữ liệu và xử lý dữ liệu tăng cường. Gán nhãn bộ dữ liệu về đối tượng trong khung hình với thông tin về tọa độ, kích thước hộp giới hạn chứa đối tượng dựa trên công cụ LabelImg. Phân chia bộ dữ liệu đã gán nhãn thông thường theo tỉ lệ 70% - tệp huấn luyện, 20% - tệp kiểm tra và 10% - tệp đánh giá.

Bước 2: Huấn luyện và đánh giá hiệu suất mô hình.

Huấn luyện mô hình là một chương trình hoạt động liên tục theo trình tự và được thực hiện trên nền tảng máy chủ ảo Google Colab [11].

Cài đặt môi trường để huấn luyện mô hình. Cấu trúc và lưu trữ các tham số quan trọng của mô hình YOLO theo file cấu hình (data.yaml) được tạo. Huấn luyện mô hình theo các dòng lệnh và theo dõi kết quả.

Đánh giá hiệu suất mô hình dựa trên các tiêu chí về độ tin cậy mô hình (precision), độ nhạy mô hình (recall), độ chính xác trung bình (mAP), giá trị mất mát (loss). Đánh giá khả năng nhận dạng mục tiêu trên tập dữ liệu theo ngưỡng xác định chỉ tiêu nhận dạng IoU_{th} .

Bước 3: Triển khai và thử nghiệm thực tế mô hình đã huấn luyện

Nền tảng triển khai mô hình và thử nghiệm là máy tính có cấu hình CPU Intel Xeon E5-2670 V3, GPU NVIDIA GeForce GTX1080. Tải và cài đặt Cuda 11.6.0, Cudnn 8.5.0 phù hợp với hệ điều hành Windows để chương trình làm việc với GPU đạt tốc độ tốt nhất. Trong môi trường lập trình C#, mô hình Yolo được tích hợp với giao diện ML.Net và ONNX Runtime.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Chuẩn bị cho thử nghiệm

KCBVV-HL cỡ 72 mm được nhóm nghiên cứu thiết kế chế tạo: Hệ bắt-bám camera số bao gồm: Khối ống phóng dùng để xạ thủ ngắm vào mục tiêu; khối xử lý trung tâm với nhiệm vụ quản lý các chế độ làm việc của toàn hệ thống, truyền, nhận và xử lý các thông tin về mục tiêu và các kênh xạ thủ, thực hiện các thuật toán bắt-bám mục tiêu trên cơ sở thông tin về hình ảnh mục tiêu do camera trên ống phóng cung cấp (hình 6).



Hình 6. KCBVV-HL cỡ 72 mm và khối xử lý trung tâm.



Hình 7. Mục tiêu treo trên cáp giữa 2 cơ cấu cột.



Hình 8. Giao diện phần mềm thử nghiệm.

Hệ camera số và thị giác máy tính có các tính năng: Trường nhìn của ống kính 26,2×19,8 (°); độ phân giải cảm biến 1920×1080 (Pixels). Với các tham số trên của camera thì kích thước toàn ảnh là 1920×1080 (Pixels) tương đương góc nhìn 26,2×19,8 (°); kích thước vòng tròn thị giới ảo là 222×222 (Pixels) tương đương góc nhìn 3×3 (°).

Theo tính năng camera được sử dụng, trường nhìn mở rộng của hệ bắt-bám camera số đảm bảo xấp xỉ $\pm 13^\circ$ theo mặt phẳng đứng và $\pm 10^\circ$ theo mặt phẳng ngang. Bài báo tập trung vào thực hiện giả lập hoạt động của thị giới đĩa điều chế ($\pm 1,5^\circ$) trong chế độ hãm và tháo hãm. Thực tế, khi KCBVV-HL còn ở trong ống phóng, trường nhìn mở rộng yêu cầu nhỏ hơn so với quá trình bay tự dẫn ngoài ống phóng, bởi trắc thủ liên tục rê ống phóng hướng vào mục tiêu, sai số rê ống phóng chỉ khoảng dưới 10° . Tuy nhiên, nâng cao về trường nhìn mở rộng sát với thực tế là nội dung cần được tiếp tục xem xét và giải quyết.

Mô hình mục tiêu (hình 7): Mô hình máy bay F18 Hornet di động dạng 3D có kích thước (450x320x90) mm, trọng lượng 1 kg (cả cơ cấu giá, treo). Kích thước mục tiêu thiết kế đảm bảo tạo khung hình ảnh mục tiêu với cự li 9 ÷ 40 mét tương ứng với khung hình lớn nhất là 220×220 Pixels và nhỏ nhất 20×20 Pixels. Mục tiêu có thể di động ngang trên cáp treo 40 m giữa hai cụm cột cố định, được điều khiển với vận tốc thay đổi 1,0 ÷ 2,5 m/s.

Giao diện phần mềm thử nghiệm của KCBVV-HL cỡ 72 mm với hệ camera số được thể hiện ở hình 8.

3.2. Kết quả huấn luyện và thử nghiệm

Thực hiện huấn luyện mô hình YOLO phiên bản 8 nano 250 vòng trên tập dữ liệu 5000 ảnh mẫu, kết quả thử nghiệm thu được các tham số của mô hình học sâu (hình 9) dựa trên các tiêu chí sau: độ tin cậy mô hình học sâu (precision), độ nhạy mô hình học sâu (recall), độ chính xác trung bình (mAP), giá trị mất mát (loss).

Kết quả nhận dạng và bắt-bám mục tiêu được trình bày trên hình 10.

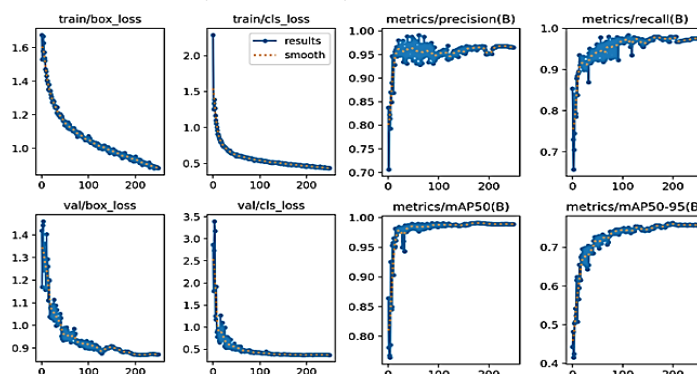
- Thảo luận:

Kết quả huấn luyện thu được các giá trị precision $\approx 0,97$; recall $\approx 0,98$ (hình 9) cho thấy mô hình học sâu đạt được sự cân bằng giữa hai tiêu chí này, giá trị precision và recall đều rất cao (gần đạt max là 1).

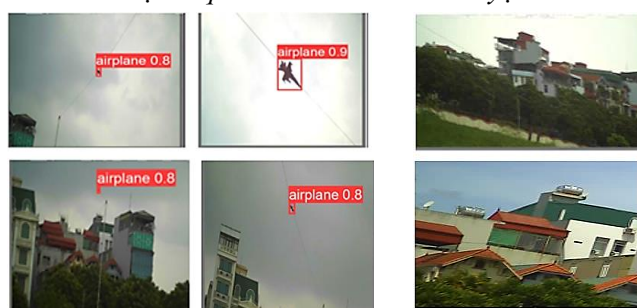
Giá trị mAP trong quá trình huấn luyện tăng dần và đạt mAP50 $\approx 0,99$; mAP50-95 $\approx 0,78$ (hình 9) cho thấy mô hình hoạt động tốt và đáp ứng yêu cầu về bắt-bám mục tiêu.

Giá trị mất mát loss trên các tập dữ liệu huấn luyện (train) và tập dữ liệu đánh giá (val) sau 100-

200 epochs huấn luyện giảm dần và đạt xấp xỉ 0,87 (hình 9). Điều này chứng tỏ mô hình đang dự đoán vị trí mục tiêu sát với nhãn thực của dữ liệu mẫu.



Hình 9. Đồ thị kết quả các tham số huấn luyện của mô hình.



a) Có mục tiêu.

b) Không có mục tiêu.

Hình 10. Kết quả nhận dạng và bắt-bám mục tiêu.

Hình 10 mô tả một số kết quả nhận dạng và bắt bám mục tiêu điển hình, với chỉ số IoU xấp xỉ 0,8 thỏa mãn yêu cầu nhận dạng mục tiêu đúng và không nhận nhầm đối tượng. Khi mục tiêu ở gần, nó được quan sát rõ nét hơn, do đó, khả năng phát hiện và nhận dạng chính xác là rất cao. Khi mục tiêu ở giới hạn xa, độ khó trong việc phát hiện sẽ tăng lên do kích thước ảnh mục tiêu nhỏ đi nhiều và khó nhận biết hơn, gây khó khăn cho quá trình nhận dạng. Điều này cũng đồng nghĩa với việc quá trình huấn luyện sẽ trở nên khó khăn hơn và kéo dài hơn, đồng thời độ chính xác cũng sẽ giảm. Tuy nhiên, vấn đề này được giải quyết bằng cách đặt giới hạn cự li chuyển động của mục tiêu (hạn xa, hạn gần) để hòa hợp các yếu tố: trường nhìn của camera, độ phân giải của camera, kích thước mục tiêu và kích thước ảnh tối thiểu của mục tiêu mà mô hình có thể nhận dạng. Qua thử nghiệm tại cự li giới hạn xa 40 m chỉ số IoU vẫn đạt tối thiểu 0,7 theo yêu cầu về ngưỡng chỉ tiêu nhận dạng.

Với điều kiện thử nghiệm ban ngày thời tiết đơn giản, theo trình tự thu thập bộ dữ liệu mục tiêu thực địa, xử lý hình ảnh và huấn luyện mô hình học sâu nhận dạng mục tiêu, thử nghiệm trên hệ camera số và máy tính cấu hình CPU Intel Xeon E5-2670 V3, GPU NVIDIA GeForce GTX1080, nhóm nghiên cứu đã thu được tốc độ xử lý ảnh của thị giác máy tính đạt 14 hình/giây và dữ liệu video về mục tiêu di động từ hệ camera số. Thử nghiệm trên một KCBVV-HL cỡ 72 mm, thuật toán học sâu YOLO phiên bản 8 đáp ứng được khả năng bắt-bám các mục tiêu với thời gian đáp ứng nhận dạng mục tiêu xấp xỉ 300 ms, phù hợp với thời gian đáp ứng của bộ tọa độ đầu tự dẫn KCBVV cỡ 72 mm trong giai đoạn ngắm bắn.

4. KẾT LUẬN

Về mặt ứng dụng, kết quả nghiên cứu và thực nghiệm của bài báo đã đề ra cơ sở ứng dụng để chế tạo các thiết bị KCBVV-HL khác nhau với tính năng tương đương. Tuy nhiên, nghiên cứu vẫn còn

nhược điểm về vấn đề nâng cao chất lượng nhận dạng mục tiêu ở những điều kiện thời tiết phức tạp hơn (mây mù, dây treo rung lắc,...), cũng như trường nhìn mở rộng trong chế độ tháo hãm của hệ bắt-bám còn hạn chế. Hướng giải quyết vấn đề đã nêu cần được cải thiện cả phần cứng và phần mềm. Về phần cứng, để nâng trường nhìn rộng có thể thay đổi camera có độ phân giải lớn và tích hợp cụm 2 đến 3 camera cùng một chức năng thu thập ảnh, hoặc sử dụng bộ gimbal phù hợp. Vấn đề này sẽ được nhóm nghiên cứu tiếp tục nghiên cứu giải quyết và công bố trong các bài báo tiếp theo.

TÀI LIỆU THAM KHẢO

- [1]. K. Kumar *et al.*, “Comparative study on object detection in visual scenes using deep learning,” World Journal of Advanced Engineering Technology and Sciences, **Vol. 12**, No. 2, pp. 045-050, (2023).
- [2]. J. Jiang *et al.*, “Research on moving object tracking technology of sports video based on deep learning algorithm,” 4th International Conference on Information Systems and Computer Aided Education, pp. 2376-2380, (2021).
- [3]. V. G. Dhanya *et al.*, “Deep learningbased computer vision approaches for smart agricultural applications,” Artificial Intelligence in Agriculture, pp. 211-229, (2022).
- [4]. M. H. Koresh, “Computer vision based traffic sign sensing for smart transport,” Journal of Innovative Image Processing, **Vol. 1**, No.1, pp. 11-19, (2019).
- [5]. L. M. Belmonte *et al.*, “Computer vision in autonomous unmanned aerial vehicles a systematic mapping study,” Applied Sciences, **Vol. 9**, No.15, pp. 3196, (2019).
- [6]. E. Shreyas *et al.*, “3D object detection and tracking methods using deep learning for computer vision applications,” International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). IEEE, (2021).
- [7]. J. Redmon *et al.*, “You only look once: Unified, real-time object detection,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, (2016).
- [8]. J. F. Henriques *et al.*, “High-speed tracking with kernelized correlation filters,” IEEE transactions on pattern analysis and machine intelligence, **Vol. 37**, No.3, pp. 583-596, (2014).
- [9]. Cục Kỹ thuật PK-KQ, “Thuyết minh kỹ thuật và hướng dẫn sử dụng phối bộ tên lửa tầm thấp A72 (9M32M),” (2003).
- [10]. T. V. Dực, N.V. Sơn, P. V. Uy, “Động học bay và nguyên lý dẫn khí cụ bay điều khiển một kênh,” NXB Khoa học Kỹ thuật (2006).
- [11]. M. Ed. Schumann, “A Book about Colab:(and Related Activities),” Printed Matter (2015).

ABSTRACT

Application of deep learning in computer vision for the target detection-tracking system on shoulder-training flight vehicles

The paper presents an application of a deep learning model in computer vision for the target detection and tracking system on shoulder-training flight vehicles. The purpose of the research is to use the digital camera system as a simulated detection and tracking system similar to the features of the seeker of shoulder-launched flight vehicles, as a basis for building shoulder-training flight vehicles. To detect and track moving targets outdoors, 3D target detection ability must meet accuracy requirements; however, when traditional image processing algorithms are applied, they are not effective. To solve this problem, the article focuses on researching and applying the YOLO deep learning model to the digital camera tracking system. Using the YOLO version 8 model, the authors tested data collection, image processing and model training as well as considered the target detection and tracking ability of a digital camera system. At the same time, a newly designed shoulder-training flight vehicle - 72 millimeters with a digital camera tracking system was tested outdoors. As a result, detecting and tracking the target by the digital camera system similar to the shoulder-launched flight vehicle - 72 millimeters in a specifically limited range condition is obtained.

Keywords: Target detection and tracking system; Digital camera system; Deep learning model; YOLO algorithm.