

MỘT PHƯƠNG PHÁP THEO DÕI ĐỐI TƯỢNG DÀI HẠN DỰA VÀO THUẬT TOÁN SIAMRPN

A LONG-TERM OBJECT TRACKING METHOD BASED ON THE SIAMRPN ALGORITHM

Lê Khánh Thành¹, Vũ Quốc Huy¹, Lê Trần Thắng¹,
Bùi Nguyên Quân², Lê Bá Tuấn^{1,*}

DOI: <https://doi.org/10.57001/huivh5804.2025.374>

TÓM TẮT

Bài báo này đề xuất một phương pháp cải tiến cho bài toán theo dõi đối tượng dựa trên nền tảng thuật toán SiamRPN kết hợp với cơ chế cập nhật mẫu có điều kiện. Phương pháp nhằm khắc phục hạn chế của các bộ theo dõi sử dụng mẫu cố định, vốn dễ mất mục tiêu trong các tình huống theo dõi dài hạn khi đặc điểm hình thái của đối tượng thay đổi đáng kể. Thuật toán sẽ chỉ cập nhật mẫu khi phát hiện sự thay đổi rõ rệt về kích thước hình học giữa ảnh mẫu ban đầu và đối tượng hiện tại, đồng thời đạt ngưỡng độ tin cậy định sẵn, giúp tăng tính ổn định và giảm hiện tượng trôi mẫu đồng thời vẫn giữ được hiệu suất và tốc độ của mô hình. Các thí nghiệm trên tập dữ liệu chuẩn OTB-2015 và tập dữ liệu ảnh nhiệt AFII25 cho thấy phương pháp đề xuất đạt độ chính xác cao, tương đương với các phương pháp hiện đại như ECO, TDA-Track, nhưng vượt trội về tốc độ xử lý (lên tới 190 FPS). Phương pháp cho thấy khả năng hoạt động hiệu quả trong các môi trường quan sát khó khăn như ảnh nhiệt, điều kiện thời tiết thay đổi và che khuất tạm thời, từ đó phù hợp với các ứng dụng theo dõi thời gian thực trên UAV, thiết bị nhúng và hệ thống giám sát thông minh.

Từ khóa: Theo dõi bắt bám, SiamRPN, OTB-2015, AFII25.

ABSTRACT

This paper proposes an improved method for object tracking based on the SiamRPN algorithm combined with a conditional template update mechanism. The method aims to overcome the limitations of trackers using fixed templates, which are prone to losing targets in long-term tracking situations when the morphological characteristics of the object change significantly. The algorithm will only update the template when it detects a clear change in geometric size between the initial template image and the current object, and at the same time reaches a predetermined confidence threshold, helping to increase stability and reduce template drift while still maintaining the performance and speed of the model. Experiments on the OTB-2015 benchmark dataset and the AFII25 thermal image dataset show that the proposed method achieves high accuracy, equivalent to modern methods such as ECO, TDA-Track, but is superior in processing speed (up to 190 FPS). The method shows the ability to operate effectively in difficult observation environments such as thermal imaging, changing weather conditions and temporary occlusion, thus suitable for real-time tracking applications on UAVs, embedded devices and smart surveillance systems.

Keywords: Object tracking, SiamRPN, OTB-2015, AFII25.

¹Viện Tự động hóa, Viện Khoa học và Công nghệ Quân sự

²Viện Nghiên cứu 486

*Email: batuanle@hotmail.com

Ngày nhận bài: 13/8/2025

Ngày nhận bài sửa sau phản biện: 18/11/2025

Ngày chấp nhận đăng: 28/11/2025

1. GIỚI THIỆU

Theo dõi bắt bám mục tiêu là một nhánh rất quan trọng trong lĩnh vực thị giác máy tính và công nghệ này được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau

như quân sự, xe tự hành và an ninh... Tuy nhiên các thuật toán bắt bám mục tiêu còn nhiều thách thức trong các ứng dụng thực tế, chẳng hạn như: sự nhiễu động của các mục tiêu giống nhau, độ biến dạng mục tiêu lớn, mục tiêu

di chuyển nhanh,... Do đó, việc cải thiện độ chính xác và hiệu suất thời gian thực của việc theo dõi mục tiêu đang được nhiều nhà khoa học đặt làm trọng tâm nghiên cứu.

Các bước thực hiện cơ bản của thuật toán theo dõi mục tiêu được mô tả như sau: Đầu tiên, trích xuất đặc điểm được thực hiện trên mẫu ứng cử viên của mục tiêu đã biết, sau đó kết hợp với thông tin, đặc điểm của các vùng ứng viên kế tiếp để đối chiếu, công việc này được thực hiện và kết quả theo dõi cuối cùng do mô hình đã đào tạo [1]. Các phương pháp theo dõi bám bắt mục tiêu truyền thống chủ yếu sử dụng các tính năng thủ công và được chia thành hai loại dựa trên các mô hình quan sát khác nhau: phương pháp mô hình tổng quát và phương pháp mô hình phân đoạn [2]. Những năm gần đây, cùng với sự ra đời và phát triển vượt bậc của học sâu. Nó đã được áp dụng mạnh mẽ trong lĩnh vực thị giác máy tính, và các thuật toán theo dõi bám bắt mục tiêu cũng không ngoại lệ. Do học sâu có khả năng trích xuất thông tin ngữ nghĩa và khả năng khái quát hóa mạnh mẽ nên học sâu dần được sử dụng rộng rãi trong lĩnh vực theo dõi và đạt được nhiều kết quả đáng ghi nhận. Theo cách hiểu và giải quyết nhiệm vụ thì thông thường phương pháp theo dõi mục tiêu dựa trên học sâu được chia thành: theo dõi mục tiêu dựa trên phân loại, theo dõi mục tiêu dựa trên hồi quy và theo dõi mục tiêu dựa trên sự kết hợp giữa phân loại và hồi quy. Marvasti-Zadeh và cộng sự [3] đã có một bài tổng quan thuật toán theo dõi mục tiêu dựa vào học sâu từ nhiều góc độ khác nhau, như cấu trúc mạng, phương pháp huấn luyện mạng, chức năng mạng và đầu ra mạng. Fiaz và cộng sự [4] đã chia các thuật toán theo dõi mục tiêu hiện tại thành hai loại, dựa trên lọc tương quan và lọc không tương quan, đồng thời phân loại và giới thiệu thêm các thuật toán theo dõi mục tiêu khác nhau trong hai loại trên theo các cấu trúc khung mạng khác nhau. Một bộ dữ liệu theo dõi mục tiêu mới OTTC được đề xuất và các thử nghiệm so sánh giữa các thuật toán khác nhau được tiến hành trên bộ dữ liệu này. Javed và cộng sự [5] thực hiện một nghiên cứu so sánh về các thuật toán theo dõi mục tiêu dựa trên lọc tương quan và mạng song sinh sâu, đồng thời phân tích đánh giá hiệu suất của mô hình này trên nhiều bộ dữ liệu. Soleimanitaleb và cộng sự [6] đã chia các thuật toán theo dõi mục tiêu hiện tại thành bốn loại dựa trên tính năng, phân đoạn, ước lượng và cách huấn luyện, tập trung giới thiệu các thuật toán theo dõi dựa trên quy trình huấn luyện.

Trong bài toán theo dõi bám bắt mục tiêu, thì theo dõi mục tiêu đơn là một nhánh riêng, có nhiều ứng dụng trong thực tế, vì thế nó cũng được nhiều nhà khoa học đào sâu nghiên cứu. Hiện nay, việc theo dõi bám bắt mục tiêu đơn vẫn còn nhiều thách thức do những khó khăn

như đối tượng biến dạng, bị che khuất, thay đổi tỷ lệ và sự phức tạp của nền. Trong vài thập kỷ qua, các nghiên cứu chủ yếu đến từ các hội nghị thị giác máy tính hàng đầu thế giới CVPR, ECCV và ICCV cùng với đó là các cuộc thi được tổ chức trên toàn thế giới. Ý tưởng thiết kế của thuật toán theo dõi mục tiêu đơn có thể được chia thành ba phần: phương pháp truyền thống dựa trên thuật toán xử lý tín hiệu hình ảnh truyền thống và học máy; phương pháp lọc tương quan với tính năng khớp mẫu, và cuối cùng là phương pháp học sâu để đào tạo mô hình sau đó sử dụng phân loại hoặc hồi quy ra đối tượng cuối. Năm 2016, thuật toán SINT và SiamFC lấy ý tưởng từ từ mạng thần kinh song sinh (Siamese Neural Network, SNN) đã cho ra các mô hình theo dõi mới dựa trên học sâu. Tiếp sau, SiamRPN ra đời và có những cải tiến vượt trội, giúp cấu trúc SNN tỏa sáng trong các tập dữ liệu chuẩn VOT2018, VOT2019 và VOT2020 [7]. Khác với các phương pháp theo dõi truyền thống thường chỉ dự đoán một bounding box duy nhất, SiamRPN sử dụng RPN để sinh ra nhiều vùng đề xuất (proposal boxes) và đánh giá mức độ tin cậy cho từng vùng đó, từ đó xác định chính xác vị trí của đối tượng. Kiến trúc này không chỉ tăng độ chính xác mà còn giữ được tốc độ xử lý rất cao. Trong các thử nghiệm, SiamRPN có thể chạy với tốc độ lên đến hơn 200 FPS, đủ để ứng dụng trong các hệ thống thời gian thực như giám sát an ninh, robot tự hành hoặc UAV trên các mạnh nhúng nhỏ gọn. Một điểm mạnh nữa của SiamRPN là tính tổng quát tốt. Do được huấn luyện trước với nhiều cặp ảnh khác nhau, nó có khả năng xử lý các biến đổi về hình dạng, ánh sáng, góc nhìn và cả trường hợp đối tượng bị che khuất một phần. Tuy nhiên, như nhiều thuật toán theo dõi đơn mục tiêu khác, SiamRPN vẫn gặp khó khăn trong các tình huống phức tạp như đối tượng bị che khuất hoàn toàn trong thời gian dài, hoặc có nhiều vật thể tương tự nhau trong cùng một khung hình. Cùng với đó là sự thay đổi liên tục về kích thước và hình dạng của đối tượng, dẫn đến khả năng mất bám với thời gian dài.

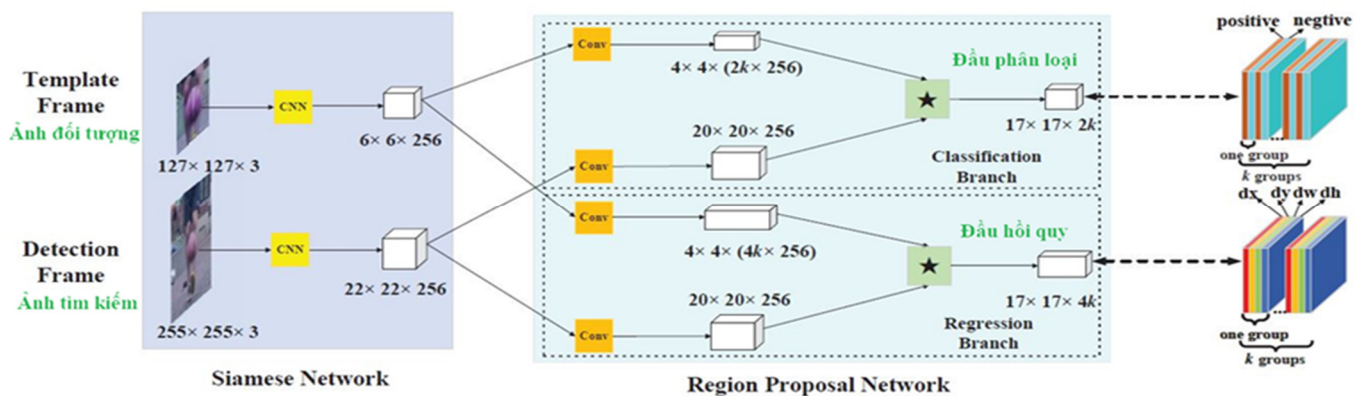
Bài báo này đưa ra một phương pháp bắt bám mục tiêu đơn dựa trên thuật toán SiamRPN. Dữ liệu được lấy từ nhiều nguồn bao gồm các tập dữ liệu chuẩn và dữ liệu tự tạo của nhóm nghiên cứu. Thông qua huấn luyện cấu trúc SiamRPN để đưa ra các mô hình bắt bám phù hợp. Một cơ chế cải tiến trong phần suy luận bắt bám được đề xuất để tăng hiệu quả cho quá trình bắt bám mục tiêu dài. Thông qua các thí nghiệm để đánh giá hiệu suất của mô hình.

2. PHƯƠNG PHÁP

2.1. Thuật toán SiamRPN

Khung tổng thể của SiamRPN được thể hiện trong hình 1, khung đề xuất bao gồm một mạng con Siamese

để trích xuất tính năng và một mạng con Region proposal kế tiếp dùng để đưa ra vùng đối tượng. Cụ thể, ở mạng Siamese có 2 nhánh con, một nhánh là ảnh đầu vào đối tượng, nhánh còn lại là ảnh tổng thể. Ở mạng Region proposal, một nhánh phụ trách phân loại tiền đối tượng - nền, nhánh còn lại được sử dụng để đưa ra khung đối tượng. Chính vì vậy, mô hình được huấn luyện từ đầu đến cuối, không qua nhiều bước. Quá trình huấn luyện sẽ lấy ảnh đầu tiên làm ảnh đối tượng, các ảnh sau được nhập liên tục vào ảnh tìm kiếm, từ đó huấn luyện ra trọng số tối ưu của mô hình.



Hình 1. Khung tổng thể của thuật toán SiamRPN

Trong quá trình suy luận của thuật toán này. Nhóm tác giả đã đưa ra hai phương pháp để lựa chọn đối tượng cuối cùng. Phương pháp đề xuất đầu tiên là loại bỏ các hộp giới hạn được tạo ra bởi các điểm neo quá xa tâm. Ví dụ, chúng ta chỉ giữ lại vùng con $g \times g$ ở tâm trên bản đồ đặc trưng phân loại. Vì thường giữa 2 khung ảnh nối tiếp nhau sẽ không có sự khác biệt lớn, nên phương pháp loại bỏ các điểm neo ở xa có thể loại bỏ hiệu quả các điểm ngoại lai. Hình 2 minh họa việc chọn các điểm neo mục tiêu có khoảng cách không quá 7 pixel từ tâm của bản đồ đặc trưng phân loại.

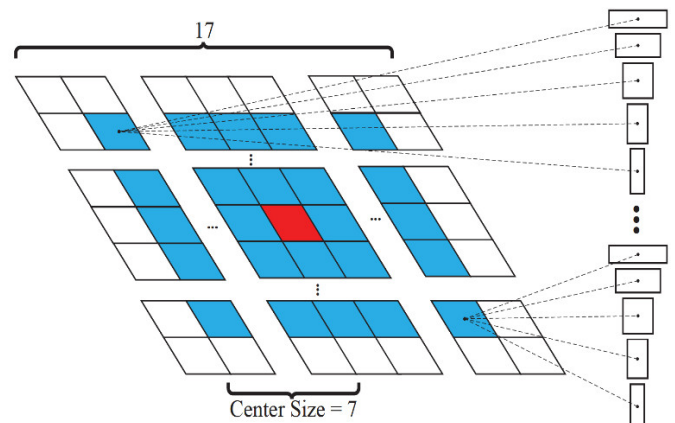
Phương pháp lựa chọn đề xuất thứ hai của nhóm tác giả là sử dụng cửa sổ cosin và hình phạt (penalty) để xếp hạng lại điểm số của các đề xuất nhằm tìm ra điểm số tốt nhất. Sau khi loại bỏ các giá trị ngoại lệ, một cửa sổ cosin được thêm vào để loại bỏ độ dịch chuyển lớn của đối tượng, và sau đó một hình phạt được thêm vào để loại bỏ sự thay đổi lớn về kích thước và tỷ lệ dựa theo công thức (1):

$$\text{penalty} = e^{k \times \max\left(\frac{r}{r'}, \frac{r'}{r}\right) \times \max\left(\frac{s}{s'}, \frac{s'}{s}\right)} \quad (1)$$

Ở đây k là siêu tham số, r biểu thị tỷ lệ chiều cao và chiều rộng của đề xuất và r' biểu thị tỷ lệ của khung hình cuối cùng. s và s' biểu thị tỷ lệ tổng thể của đề xuất và khung hình cuối cùng, được tính như công thức (2):

$$(w + p) \times (h + p) = s^2 \quad (2)$$

trong đó, w và h là chiều rộng và chiều cao của mục tiêu, p được tính theo công thức $(w+h)/2$. Sau các bước trên, các đề xuất K hàng đầu sẽ được xếp hạng lại sau khi nhân điểm phân loại với điểm hình phạt. Phương pháp loại trừ không tối đa (Non-maximum-suppression, NMS) được thực hiện sau đó để có được khung giới hạn theo dõi cuối cùng. Sau khi khung được chọn, kích thước mục tiêu được cập nhật bằng nội suy tuyến tính để giữ cho hình dạng thay đổi mượt mà hơn.



Hình 2. Phương pháp loại các đối tượng xa tâm neo

2.2. Đề xuất 1 thuật toán học trực tuyến

Thuật toán SiamRPN là một trong những phương pháp theo dõi đối tượng một cách trực tiếp, dựa trên việc sử dụng ảnh mẫu (template) lấy từ khung hình đầu tiên để thực hiện so khớp với các khung hình tiếp theo. Tuy nhiên, trong các kịch bản theo dõi dài hạn, đặc điểm hình thái, kích thước, hoặc góc nhìn của đối tượng thường thay đổi đáng kể so với ảnh mẫu ban đầu, dẫn đến hiện tượng suy giảm độ chính xác trong quá trình theo dõi.

Để khắc phục nhược điểm này và cũng đảm bảo tốc độ xử lý của thuật toán, chúng tôi đề xuất một cơ chế cập nhật có điều kiện ảnh mẫu, dựa trên sự thay đổi kích thước hình học của đối tượng và độ tin cậy của bộ theo

đổi. Cụ thể, sau mỗi chuỗi khung hình theo dõi, nếu tỉ lệ thay đổi kích thước giữa đối tượng hiện tại và ảnh mẫu ban đầu vượt qua một ngưỡng xác định trước, đồng thời độ tin cậy của bộ theo dõi đạt trên một giá trị ngưỡng cụ thể thì ảnh mẫu ban đầu sẽ được cập nhật bằng ảnh của đối tượng hiện tại. Việc cập nhật này nhằm mục tiêu là đảm bảo rằng ảnh mẫu luôn phản ánh tương đối chính xác hình thái hiện tại của đối tượng, từ đó duy trì hiệu quả theo dõi ổn định trong thời gian dài. Phương pháp đề xuất được mô tả như sau: Tại thời điểm t_0 chúng ta chọn ảnh mẫu z_0 ; tại mỗi thời điểm t_i , có đầu ra khung B_i và điểm tin cậy là s_i . Ta có kích thước của đối tượng tại thời điểm t_0 và t_i lần lượt là A_0 và A_i , được tính theo công thức (3).

$$\begin{cases} A_0 = w_0 \times h_0 \\ A_i = w_i \times h_i \end{cases} \quad (3)$$

ở đây: w_0 , h_0 và w_i , h_i lần lượt là kích thước của khung tại thời điểm t_0 và t_i . Chúng ta dựa theo công thức (4) để cập nhật ảnh mẫu z_0 .

$$\begin{cases} \left| \frac{A_i - A_0}{A_0} \right| > \theta_A \\ s_i > \theta_s \end{cases} \quad (4)$$

ở đây, θ_A và θ_s là ngưỡng thay đổi kích thước và ngưỡng điểm tin cậy. Phương pháp này giữ cho mẫu đối tượng được cập nhật thích nghi nhưng có kiểm soát, tránh việc cập nhật sai trong các tình huống đối tượng bị che khuất hoặc theo dõi nhầm đối tượng tương tự. Chúng tôi gọi phương pháp đề xuất này là học trực tuyến nhanh SiamRPN (FOL-SiamRPN).

3. KẾT QUẢ VÀ THẢO LUẬN

Để đánh giá hiệu suất của thuật toán nhóm nghiên cứu đề xuất, chúng tôi sử dụng tập dữ liệu chuẩn OTB-2015 [8] và tập dữ liệu do nhóm nghiên cứu tự tạo chúng tôi gọi là AFII25. Sử dụng các chỉ số đánh giá chuẩn thường dùng, bao gồm overlap precision (OP): đo lường tỉ lệ khung dự đoán có độ giao nhau (IoU) $\geq 0,5$ so với khung thực; distance precision (DP): đo tỉ lệ khung dự đoán nằm trong bán kính 20 pixel tính từ vị trí trung tâm của khung thực [9] và tốc độ xử lý khung hình trên 1 giây (FPS). Chúng tôi sử dụng một số phương pháp bắt bám mục tiêu tiên tiến hiện nay để so sánh và đánh giá với phương pháp được đề xuất của chúng tôi, bao gồm BACF [10], ECO [11], TDA-Track [12]. Kết quả so sánh được chạy trên phần mềm hệ thống Window 10, ngôn ngữ lập trình python, với phần cứng là máy tính sử dụng card đồ họa NVIDIA RTX 3060 12G VRAM.

3.1. Trên tập dữ liệu OTB-2015

Thông qua các thử nghiệm trên tập dữ liệu OTB-2015 với các siêu tham số θ_A và θ_s cho phương pháp FOL-SiamRPN. Chúng tôi đưa ra sự lựa chọn tối ưu với ngưỡng cập nhật mẫu $\theta_A \geq 1$ và $\theta_s \geq 0,25$. Dựa vào bảng 1, có thể thấy rằng phương pháp ECO đạt độ chính xác cao nhất về chỉ số OP và DP, tuy nhiên lại gặp phải hạn chế lớn về thời gian xử lý, chỉ đạt 12 FPS. Trong khi đó, TDA-Track thể hiện hiệu năng khá ổn định với độ chính xác trên 80% và tốc độ xử lý chấp nhận được. Ngược lại, BACF có cả độ chính xác lẫn tốc độ xử lý đều tương đối thấp. Đối chiếu với các phương pháp trên, thuật toán mà chúng tôi đề xuất tuy có độ chính xác chỉ thấp hơn không đáng kể so với ECO và TDA-Track, nhưng lại nổi bật với tốc độ xử lý vượt trội lên đến 190 FPS. Từ những so sánh này, có thể khẳng định rằng phương pháp của chúng tôi đã đạt được sự cân bằng hợp lý giữa độ chính xác và hiệu suất thời gian. Với khả năng xử lý nhanh gấp nhiều lần các phương pháp khác, thuật toán này đặc biệt phù hợp cho các ứng dụng yêu cầu phản hồi tức thời như theo dõi đối tượng trên drone, robot tự hành, hoặc các hệ thống giám sát thông minh hoạt động trong môi trường tài nguyên hạn chế. Đây cũng chính là một trong những lý do chính để chúng tôi lựa chọn thuật toán này thay vì các thuật toán tiên tiến khác như SiamMask, SiamBAN, hay TransT.

Bảng 1. So sánh độ chính xác giữa các phương pháp trên tập OTB-2015

Phương pháp	OP (%)	DP (%)	FPS
BACF	76,7	81,5	40
ECO	84,9	91,0	12
TDA-Track	84	88,1	80
FOL-SiamRPN	82,3	85,8	190

Ngoài ra, một lợi thế khác là tính ổn định của phương pháp khi xử lý các tình huống phức tạp, chẳng hạn như đối tượng di chuyển nhanh hoặc xuất hiện nhiễu động trong khung hình. Nhờ sự tối ưu hóa về mặt thuật toán và khai thác hiệu quả tài nguyên phần cứng, hệ thống không chỉ giảm thiểu độ trễ mà còn góp phần tiết kiệm năng lượng trong quá trình vận hành. Tóm lại, phương pháp đề xuất của chúng tôi không chỉ là giải pháp cân bằng giữa độ chính xác và tốc độ, mà còn chứng minh được tính khả thi trong thực tế, nhất là với các hệ thống yêu cầu hiệu năng cao và tính ổn định dài hạn.

3.2. Trên tập dữ liệu AFII25

Tập dữ liệu AFII25 được chụp từ camera ảnh nhiệt hồng ngoại các máy bay dân sự, gồm 8 đoạn video có độ dài từ 250 tới 2120 ảnh. Các máy bay này được quay chụp với các điều kiện thời tiết khác nhau và với nhiều hướng

riêng biệt, chuyên dùng để thử nghiệm các thuật toán bắt bám mục tiêu đơn.

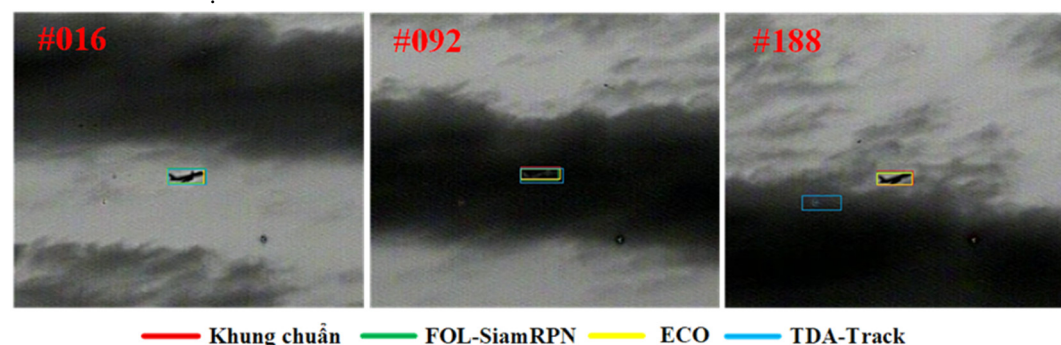
Bảng 2. So sánh độ chính xác giữa các phương pháp trên tập AFII25

Phương pháp	OP (%)	DP (%)
ECO	81,1	84,0
TDA-Track	79,3	83,2
FOL-SiamRPN	81,0	83,8

Trong bảng 2, chúng tôi đưa ra kết quả so sánh độ chính xác giữa các phương pháp ECO, TDA-Track và phương pháp cải tiến FOL-SiamRPN do chúng tôi đề xuất trên tập dữ liệu AFII25. Kết quả cho thấy rằng FOL-SiamRPN đạt độ chính xác OP là 81,0% và DP là 83,8%, gần tương đương với phương pháp ECO (OP: 81,1%, DP: 84,0%) vốn được xem là một trong những phương pháp có độ chính xác rất cao trong theo dõi mục tiêu ngắn và trung hạn.

Tuy nhiên, một điểm khác biệt đáng chú ý nằm ở tốc độ xử lý khung hình (FPS) của các phương pháp. Trong khi ECO hoạt động với tốc độ chỉ khoảng 12 FPS và TDA-Track đạt khoảng 80 FPS, thì FOL-SiamRPN cho tốc độ trung bình lên tới 190 FPS trên cùng một cấu hình phần cứng. Điều này khẳng định rằng phương pháp đề xuất không chỉ đảm bảo độ chính xác bắt bám cao, mà còn đáp ứng tốt yêu cầu xử lý thời gian thực, một yếu tố then chốt trong các hệ thống giám sát, theo dõi tự động và ứng dụng thực địa ngoài trời.

Hiệu quả của FOL-SiamRPN phần lớn đến từ chiến lược cập nhật mẫu có điều kiện, cho phép thuật toán thích nghi với sự thay đổi hình thái của đối tượng trong quá trình theo dõi dài hạn mà không cần can thiệp học lại hay tăng thêm độ phức tạp mô hình. Cụ thể, bằng cách so sánh tỉ lệ thay đổi kích thước giữa ảnh mẫu ban đầu và đối tượng hiện tại, kết hợp với ngưỡng độ tin cậy, thuật toán sẽ chỉ cập nhật mẫu khi thực sự cần thiết. Điều này giúp tránh được các cập nhật sai lệch trong trường hợp đối tượng bị che khuất, quay lưng, hoặc xảy ra nhiễu ảnh nhiệt.



Hình 3. So sánh khả năng bắt bám mục tiêu thực tế của các thuật toán

Thực tế trên tập dữ liệu AFII25, được quay bằng camera ảnh nhiệt với điều kiện thời tiết và góc nhìn đa dạng, cho thấy rõ ưu điểm của chiến lược cập nhật mẫu thích nghi này. Trong các đoạn video dài tới hơn 2000 khung hình, FOL-SiamRPN vẫn duy trì được độ chính xác cao và ổn định theo thời gian, trong khi các phương pháp truyền thống như ECO dễ gặp hiện tượng trôi mẫu hoặc đánh mất mục tiêu sau khi đối tượng có những thay đổi tức thì.

Trong hình 3, chúng ta có thể thấy rằng phương pháp TDA-Track đã gặp khó khăn khi đối tượng di chuyển qua vùng có mây dày đặc, dẫn đến hiện tượng mất mục tiêu và không thể tái xác định vị trí chính xác của đối tượng trong các khung hình tiếp theo. Điều này phản ánh điểm yếu của TDA-Track trong việc xử lý các tình huống có che khuất tạm thời hoặc thay đổi mạnh về đặc trưng thị giác.

Ngược lại, hai phương pháp còn lại là ECO và FOL-SiamRPN vẫn duy trì khả năng bắt bám ổn định xuyên suốt chuỗi khung hình, cho thấy tính bền vững trong quá trình theo dõi. Đặc biệt, FOL-SiamRPN thể hiện rõ lợi thế với cơ chế cập nhật mẫu có điều kiện, giúp thuật toán nhanh chóng thích nghi với sự thay đổi ngoại cảnh và đặc điểm đối tượng, từ đó duy trì được vùng theo dõi chính xác ngay cả khi tín hiệu ảnh nhiệt bị suy giảm tạm thời. Điều này khẳng định tính hiệu quả của phương pháp đề xuất trong các kịch bản quan sát ngoài trời có điều kiện bất lợi như sương mù, mây che, hoặc chênh lệch nhiệt độ mạnh.

Tóm lại, phương pháp FOL-SiamRPN là một hướng cải tiến đơn giản nhưng hiệu quả cho bài toán theo dõi mục tiêu trong thời gian dài, đặc biệt trong các môi trường khó khăn như ảnh nhiệt hồng ngoại, nơi thông tin hình ảnh bị suy giảm mạnh và tính nhất quán của đối tượng dễ bị phá vỡ. Phương pháp này đồng thời mở ra khả năng tích hợp vào các hệ thống theo dõi trên thiết bị nhúng, UAV, hoặc các trạm quan sát lưu động nhờ hiệu suất xử lý cao và khả năng hoạt động ổn định mà không cần mô hình học sâu

phức tạp. Trong các nghiên cứu tương lai, chúng tôi sẽ tiếp tục hướng tới việc cải thiện độ chính xác, tối ưu hóa mô hình, đồng thời tích hợp thêm các cơ chế học thích nghi nhằm nâng cao khả năng theo dõi trong các kịch bản ngày càng

phức tạp hơn. Ngoài ra, hiện tại tập dữ liệu AFII25 mới dừng lại ở con số hơn 2000 bức ảnh, trong tương lai chúng tôi sẽ phát triển tập dữ liệu thành cực dài hạn, chuỗi dữ liệu có thể lớn hơn trên 10000 bức ảnh nhằm phục vụ mục đích huấn luyện và đánh giá khách quan hơn cho các mô hình bắt bám đối tượng.

4. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất phương pháp FOL-SiamRPN, là một cải tiến của SiamRPN với cơ chế cập nhật mẫu có điều kiện, nhằm nâng cao hiệu quả theo dõi mục tiêu dài hạn. Bằng cách sử dụng tiêu chí thay đổi kích thước và ngưỡng độ tin cậy, thuật toán giúp duy trì độ chính xác ổn định mà không làm tăng độ phức tạp mô hình. Kết quả thực nghiệm trên hai tập dữ liệu OTB-2015 và AFII25 cho thấy FOL-SiamRPN đạt độ chính xác tương đương các phương pháp hiện đại như ECO, TDA-Track, nhưng vượt trội về tốc độ xử lý (190 FPS), phù hợp cho các ứng dụng thời gian thực. Phương pháp cũng cho thấy tính ổn định cao trong các môi trường khó khăn như ảnh nhiễu. Trong tương lai, chúng tôi hướng tới việc tích hợp học thích nghi để tăng cường độ chính xác trong các kịch bản phức tạp hơn.

TÀI LIỆU THAM KHẢO

- [1]. Li X., et al., "Survey of visual object tracking algorithms based on deep learning," *Journal of Image and Graphics*, 24(12), 2057-2080, 2019.
- [2]. Kwon J., Lee K. M., "Tracking by sampling and integrating multiple trackers," *IEEE transactions on pattern analysis and machine intelligence*, 36(7), 1428-1441, 2013.
- [3]. Marvasti-Zadeh S. M., Cheng L., Ghanei-Yakhdan H., Kasaei S., "Deep learning for visual tracking: A comprehensive survey," *IEEE Transactions on Intelligent Transportation Systems*, 23(5), 3943-3968, 2021.
- [4]. Fiaz M., Mahmood A., Javed S., Jung S. K., "Handcrafted and deep trackers: Recent visual object tracking approaches and trends," *ACM Computing Surveys (CSUR)*, 52(2), 1-44, 2019.
- [5]. Javed S., Danelljan M., Khan F. S., Khan M. H., Felsberg M., Matas J., "Visual object tracking with discriminative filters and siamese networks: a survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5), 6552-6574, 2022.
- [6]. Soleimanitaleb Z., Keyvanrad M. A. "Single object tracking: A survey of methods, datasets, and evaluation metrics," *arXiv preprint arXiv:2201.13066*, 2022.
- [7]. Li B., Yan J., Wu W., Zhu Z., Hu X., "High performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8971-8980, 2018.

[8]. Danelljan M., Hager G., Shahbaz Khan F., Felsberg M., "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE international conference on computer vision*, 4310-4318, 2015.

[9]. Zhu Z., Wang Q., Li B., Wu W., Yan J., Hu W., "Distractor-aware siamese networks for visual object tracking," in *Proceedings of the European conference on computer vision (ECCV)*, 101-117, 2018.

[10]. Kiani Galoogahi H., Fagg A., Lucey S., "Learning background-aware correlation filters for visual tracking," in *Proceedings of the IEEE international conference on computer vision*, 1135-1143, 2017.

[11]. Danelljan M., Bhat G., Shahbaz Khan F., Felsberg M., "Eco: Efficient convolution operators for tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6638-6646, 2017.

[12]. Fu C., Wang Y., Yao L., Zheng G., Zuo H., Pan J., "Prompt-Driven Temporal Domain Adaptation for Nighttime UAV Tracking," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 9706-9713, IEEE, 2024.

AUTHORS INFORMATION

**Le Khanh Thanh¹, Vu Quoc Huy¹, Le Tran Thang¹,
Bui Nguyen Quan², Le Ba Tuan¹**

¹Institute of Automation, Academy of Military Science and Technology, Vietnam

²486 Research Institute, Hanoi, Vietnam