

CHIẾN LƯỢC DẪN ĐƯỜNG RÔ BỐT DẠNG BÁNH DỤA TRÊN GIẢI THUẬT BIT* VÀ PHƯƠNG PHÁP HỌC TĂNG CƯỜNG

WHEELED MOBILE ROBOT NAVIGATION BASED ON BIT* ALGORITHM AND REINFORCEMENT LEARNING METHOD

Vũ Ngọc Phúc¹, Bùi Như Nghĩa², Đặng Thái Việt^{2,*}

¹Trường Cao đẳng Nghề Bách Khoa Hà Nội

²Trường Cơ khí, Đại học Bách Khoa Hà Nội

*Email: viet.dangthai@hust.edu.vn

TÓM TẮT

Rô bốt di động ngày càng được sử dụng rộng rãi trong rất nhiều lĩnh vực trong đời sống và sản xuất công nghiệp. Dựa vào hệ nhận thức của rô bốt trong môi trường di chuyển, các vấn đề về dẫn đường tối ưu trong môi trường không biết trước là một trong các trở ngại trong bài toán điều khiển và dẫn đường rô bốt tự hành. Để giải quyết vấn đề nêu trên, nhóm tác giả đề xuất chiến lược dẫn đường kiểu lai giữa thuật toán BIT* cải tiến và mô hình học tăng cường (RL). Thuật toán BIT*-RL sử dụng tìm đường toàn cục trong môi trường không biết trước sử dụng thuật toán BIT*. Sau đó, mô hình học tăng cường sử dụng hai cấu trúc MobileNetV2 cho tối ưu thông số huấn luyện và mô hình UNet cho tối ưu hành động thực thi. Cuối cùng mô hình rô bốt di động ba bánh được kiểm chứng thông qua thực nghiệm và mô phỏng tính toán đã chứng minh các ưu việt trong tối ưu quá trình tìm kiếm đường đi, đáp ứng đồng thời đa nhiệm vụ.

Từ khóa: Hệ nhận thức rô bốt; Mô hình học tăng cường; Rô bốt di động; Thuật toán dẫn đường BIT*.

ABSTRACT

Mobile robots are increasingly used in many areas of life and industrial production. Based on the robot's perception system in the moving environment, the problem of optimal navigation in an unknown environment is one of the obstacles in the problem of controlling and guiding autonomous robots. To solve the above problem, the authors propose a hybrid navigation strategy between the improved BIT* algorithm and the reinforcement learning (RL) model. The BIT*-RL algorithm uses global path finding in an unknown environment using the BIT* algorithm. Then, the reinforcement learning model uses two MobileNetV2 structures for optimizing training parameters and the UNet model for optimizing execution actions. Finally, the three-wheeled mobile robot model is verified through experiments and computational simulations, demonstrating its superiority in optimizing the path finding process, meeting multiple tasks simultaneously.

Keywords: Mobile robot perception systems; Reinforcement learning model; Mobile robots; BIT* path planning.

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, rô bốt di động đã trở nên nổi bật đáng kể trong lĩnh vực tự động hóa và rô bốt, thực hiện đồng thời nhiều nhiệm vụ điều hướng môi trường sản xuất, quân sự và đời sống. Lập kế hoạch hành động và đường đi là các thành phần cơ bản trong điều hướng tự động rô bốt di động có bánh xe (WMR) [1]. Dựa vào hệ nhận thức của rô bốt về môi trường hoạt động, rô bốt phải xây dựng một quỹ đạo khả thi từ vị trí ban đầu đến đích mong muốn. Các nghiên cứu trước đây đã chỉ ra hệ cảm biến, Lidar và máy ảnh số thu thập dữ liệu từ môi trường nhằm mục đích xây dựng hệ tri thức của WMRs. Các phương pháp điều khiển và dẫn đường truyền thống thường chỉ hiệu quả khi biết vị trí của rô bốt, vật cản và môi trường đã biết trước. Sau khi xây dựng bản đồ số từ môi trường đã biết, WMR phải xây dựng chiến lược dẫn đường khả thi từ vị trí ban đầu đến đích. Hành trình này phải tuân thủ các tiêu chí cụ thể, bao gồm giảm chi phí vận hành, xác định tuyến đường thuận tiện nhất hoặc giảm thiểu thời gian di chuyển [2].

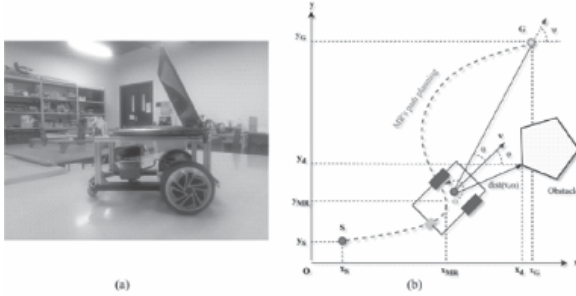
Các kế hoạch dẫn đường được phân loại thành hai loại chính: Lập kế hoạch đường đi toàn cục và cục bộ. Lập kế hoạch đường đi toàn cục (GPP) xác định một chuỗi các điểm dừng quan trọng kết nối điểm bắt đầu với điểm cuối trong tổng thể môi trường rô bốt di chuyển [3]. GPP sử dụng ba kỹ thuật cơ bản: Tìm kiếm đồ thị, tìm kiếm lấy mẫu và tìm kiếm động. Các thuật toán tìm kiếm đồ thị, chẳng hạn như Dijkstra [4] và A*[5], vượt trội trong không gian chiều thấp, đảm bảo khám phá toàn diện các tuyến đường tiềm năng. Ngược lại, các thuật toán tìm kiếm lấy mẫu phù hợp với

không gian chiều cao, cung cấp sự đảm bảo xác suất về tìm đường. Mặt khác, tìm kiếm động cải thiện khả năng kết nối của các nút đường dẫn, mặc dù phải đánh đổi bằng một số tính toán chính. Tuy nhiên, lập kế hoạch đường dẫn cục bộ (LPP) đòi hỏi khả năng tính toán nhanh, chất lượng đảm bảo, tạo ra các quỹ đạo chính xác từ nút bắt đầu đến nút mục tiêu trong một khu vực cục bộ [6]. LPP bao gồm các phương pháp truyền thống như dải đàn hồi thời gian (TEB), phương pháp cửa sổ động (DWA), và trường thế năng nhân tạo (APF) hay mạng trí tuệ nhân tạo [7].

Dựa vào các phân tích đã nêu, các chiến lược dẫn đường khác nhau có những ưu điểm và nhược điểm riêng biệt. Để tạo điều kiện cho chuyển động liền mạch và tăng cường tính ổn định của WMR bám quỹ đạo, nhóm tác giả đề xuất chiến lược dẫn đường kiểu lai kết hợp thuật toán BIT* cải tiến với mô hình học tăng cường (RL) [8]. Đầu tiên, thuật toán tìm đường BIT* được cải tiến nhằm giảm thiểu các bước tính dư thừa trong quá trình tìm đường khả thi từ điểm xuất phát đến điểm đích. Để tăng hiệu quả của quá trình lấy mẫu trong giải thuật tìm đường BIT* trong các môi trường chưa biết hoặc phức tạp, các tác giả đã kết hợp BIT* vào trong mô hình học tăng cường. Thuật toán kiểu lai BIT*-RL cung cấp những ưu điểm, bao gồm việc giảm thiểu việc lấy mẫu không hiệu quả, trích xuất các đặc điểm bản đồ kết hợp dữ liệu đã được đào tạo vào quy trình tìm đường và tăng tốc thời gian xử lý do nhu cầu xem xét mẫu giảm. Cuối cùng, các kết quả mô phỏng và thực nghiệm minh chứng sự thành công và hiệu quả cao khi áp dụng giải thuật tìm đường BIT*-RL cho điều hướng WMRs trong các môi trường chưa biết hoặc phức tạp. 

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mô hình toán học WMR



Hình 1. Rô bốt di động ba bánh di chuyển theo chiến lược dẫn đường với (a): Mô hình thực tế và (b): Biểu diễn mô hình toán học.

Mô hình rô bốt di động ba bánh sử dụng trong thực nghiệm như trong Hình 1. Dựa trên phương trình Euler-Lagrange (1) mô tả phương trình Routh động lực học phi toàn thể:

$$A(q)\ddot{q} + C(q, \dot{q})\dot{q} = E(q)\tau + F^T(q)\lambda + \Delta\zeta \quad (1)$$

Trong đó: $A(q) \in R^{n \times n}$ là ma trận đối xứng, $\tau \in R^r$ là vectơ đầu vào, $E(q) \in R^{n \times r}$ là ma trận chuyển tiếp của vectơ đầu vào, $F^T(q) \in R^{n \times n}$ là ma trận về ràng buộc. $C(q, \dot{q}) \in R^n$ biểu diễn ma trận Coriolis. $\lambda \in R^m$ là hệ số nhân Lagrange và $\Delta\zeta \in R^n$ là độ bất định trong mô hình danh nghĩa. $E(q)$ được định nghĩa trong phương trình (2) sau:

$$E(q) = \begin{bmatrix} 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}^T \quad (2)$$

Dựa trên kết quả nghiên cứu của nhóm tác giả đã công bố tại [1], mô hình toán học của WMR được xây dựng như phương trình (3):

$$\begin{cases} \ddot{\alpha} = \frac{ML_1 \sin \alpha [K_2 g - ML_1 \dot{\alpha}^2 \cos \alpha]}{K_2 K_3 - M^2 L_1^2 \cos^2 \alpha} + g_{11}(\tau_r + \tau_l) + \Delta_4 \zeta \\ \dot{v} = \frac{ML_1 \sin \alpha [K_2 \dot{\alpha}^2 - ML_1 g \cos \alpha]}{K_2 K_3 - M^2 L_1^2 \cos^2 \alpha} + g_{21}(\tau_r + \tau_l) + \Delta_5 \zeta \end{cases} \quad (3)$$

2.2. Chiến lược dẫn đường BIT*-RL

2.2.1. Thuật toán dẫn đường BIT*

Thuật toán BIT* giải quyết hiệu quả các vấn đề lập kế hoạch đường đi tối ưu tiệm cận với các điểm lấy mẫu tối thiểu trong môi trường có các chướng ngại vật đơn giản, thưa thớt. Tuy nhiên, hiệu suất của thuật toán này giảm đáng kể trong các môi trường phức tạp nơi cần nhiều điểm lấy mẫu hơn đáng kể. Khi áp dụng vào các ứng dụng thực tế, thuật toán BIT* truyền thống thể hiện hiệu suất không tối ưu do tạo ra các nút vi phạm các ràng buộc về mặt động học của rô bốt di động, dẫn đến hiệu quả tính toán kém. Hơn nữa, thuật toán này đòi hỏi phải có hậu xử lý bổ sung để lọc các nút khả thi về mặt động học trước khi xây dựng đường dẫn giải pháp cuối cùng. Nhóm tác giả đã cải tiến thuật toán dẫn đường BIT* dựa trên mô hình động lực học của WMR như trong thuật toán 1 trong Hình 2.

Algorithm 1: BIT* ($x_{start} \in X_{free}, x_{goal} \in X_{goal}$)

```

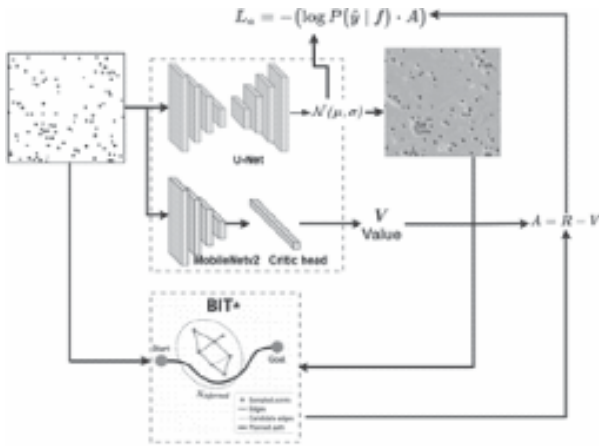
1:  $V \leftarrow \{x_{start}\}$ ;  $E \leftarrow \emptyset$ ;  $X_{samples} \leftarrow \{x_{goal}\}$ 
2:  $Q_E \leftarrow \emptyset$ ;  $Q_V \leftarrow \emptyset$ ;  $r \leftarrow \infty$ 
3: repeat
4:   if  $Q_E = \emptyset$  and  $Q_V = \emptyset$  then
5:     EscapeStrategy()
6:     Prune( $gT(x_{goal})$ );
7:      $X_{samples} \leftarrow \text{Sample}(m, gT(x_{goal}))$ ;
8:      $V_{old} \leftarrow V$ ;
9:      $V \leftarrow V \cup X_{samples}$ ;
10:     $Q_V \leftarrow V$ ;
11:     $r \leftarrow \text{radius}(|V|) + |X_{samples}|$ ;
12:   end if
13:   while Best  $Q_V \leq$  Best  $Q_E$  do
14:     ExpandVertex(BestInQueue( $Q_V$ ));
15:   end while
16:    $(v_m, x_m) \leftarrow \text{BestInQueue}(Q_E)$ ;
17:    $Q_E \leftarrow Q_E \setminus \{(v_m, x_m)\}$ ;
18:   if  $gT(v_m) + \hat{c}(v_m, x_m) + h(x_m) \leq gT(x_{goal})$  then
19:     if  $gT(v_m) + c(v_m, x_m) + h(x_m) \leq gT(x_{goal})$  then
20:       if  $gT(v_m) + c(v_m, x_m) \leq gT(V)$  then
21:          $E \leftarrow \{(v_m, x_m) \in E\}$ ;
22:       else
23:          $X_{samples} \leftarrow \{x_m\}$ ;
24:          $V \leftarrow \{x_m\}$ ;  $Q_V \leftarrow \{x_m\}$ ;
25:       end if
26:     end if
27:   else
28:      $E \leftarrow \{(v, x_m)\}$ ;
29:   end if
30:    $Q_E \leftarrow \{(v, x_m) \in Q_E \mid gT(v) + \hat{c}(v, x_m) \leq gT(x_m)\}$ ;
31:   else
32:      $Q_E \leftarrow \emptyset$ ;  $Q_V \leftarrow \emptyset$ ;
33:   end if
34: until STOP
35: return T;

```

Hình 2. Thuật toán tìm đường BIT* cải tiến

2.2.2. Mô hình học tăng cường (RL)

Thách thức chính đã nêu ở trên là việc ước tính hàm tìm kiếm BIT* với số lần lặp giới hạn, sẽ khó khăn khi xác định đường đi khả thi trong môi trường không biết. Mặc dù hàm BIT* có thể xác định đường dẫn chuyển động tối ưu, nhưng quá trình lấy mẫu của nó lại không hiệu quả vì nó không tận dụng bất kỳ thông tin nào liên quan đến môi trường và không rút ra được hiểu biết từ các vấn đề đã giải quyết trước đó. Để giải quyết những hạn chế này, các tác giả đã kết hợp mô hình học tăng cường vào thuật toán BIT*. Sự tích hợp này bao gồm một Actor xác định các hành động phù hợp cần thực hiện, trong khi Critic đánh giá hiệu quả của các hành động này và cung cấp phản hồi để điều chỉnh cần thiết. Kiến trúc tổng thể của khuôn khổ nâng cao này được mô tả trong Hình 3.



Hình 3. Cấu trúc thuật toán kiểu lai BIT*-RL

Phương pháp tính toán được xây dựng dựa trên phân phối chuẩn, được ký hiệu là $N(\mu, \sigma^2)$, trong đó:

$$\mu, \sigma^2 = \text{Actor}(\text{map}) \in \mathbb{R}^{H \times W} \quad (4)$$

Trong phương trình (4), các tham số của phân phối chuẩn được xử lý thông qua hàm Sigmoid, tạo ra một phạm vi giá trị từ 0 đến 1,

do đó tạo ra một bản đồ xác suất. Sau đó, hàm phân thưởng được xác định bằng công thức sau:

$$\text{Reward} = 1000 - \text{Cost} \quad (5)$$

Giá trị phân thưởng bằng 0 biểu thị sự vắng mặt của một đường dẫn thỏa đáng trong môi trường. Mô hình bao gồm hai thành phần chính: Actor và Critic. Actor được thiết kế bằng kiến trúc U-Net, chịu trách nhiệm tạo ra các bản đồ xác suất hoặc nói cách khác là tạo ra các hành động dựa trên đầu vào đã cho. Ngược lại, Critic sử dụng kiến trúc MobileNetV2 để đánh giá chính sách hiện tại hoặc các mẫu trọng số do Actor (π) tạo ra liên quan đến việc đạt được tham số phân thưởng. Cấu trúc kép này nâng cao khả năng và hiệu quả của mô hình trong việc tìm đường trong môi trường. Quá trình đánh giá được tiến hành thông qua hàm giá trị, trong phương trình (6):

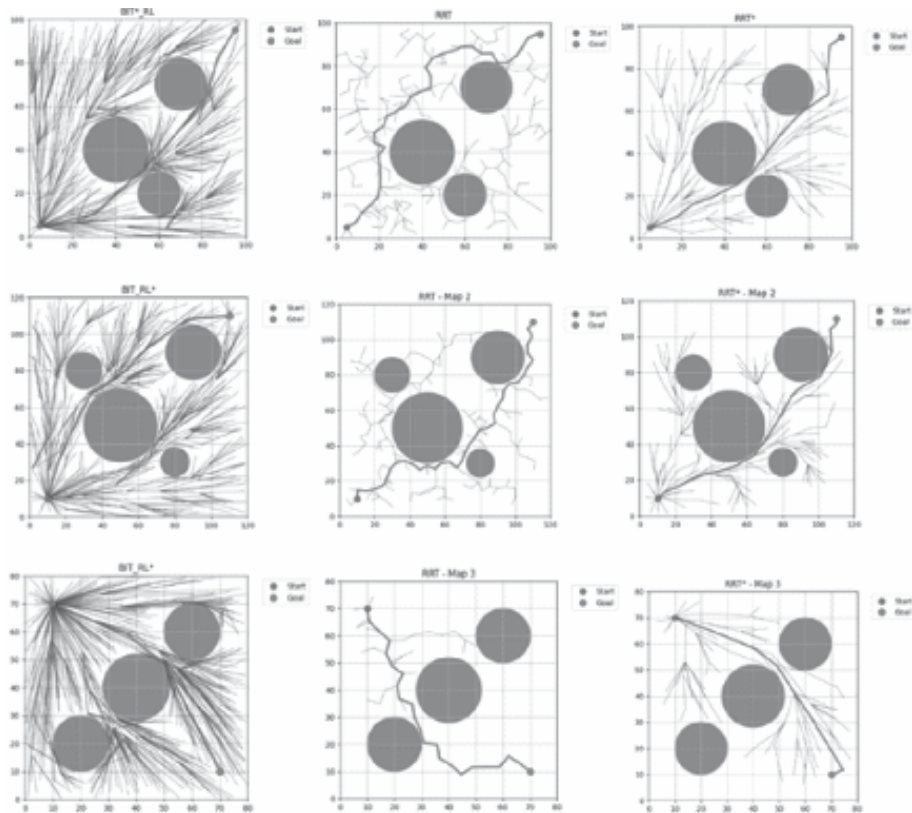
$$V(s) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right] \quad (6)$$

Trong đó: $V(s)$ là hàm giá trị tại trạng thái tương ứng s ; E_{π} là kỳ vọng trung bình theo chính sách π được ánh xạ từ trạng thái; $\gamma \in [0, 1]$ là hệ số chiết khấu; Phần thưởng tại trạng thái t được ký hiệu là r_t . Hàm $V(t)$ được tính toán để so sánh các phần thưởng thực tế, chọn các lợi thế (A) và tham gia tính toán tổn thất của Critic:

$$A = \text{Reward} - V(t) \quad (7)$$



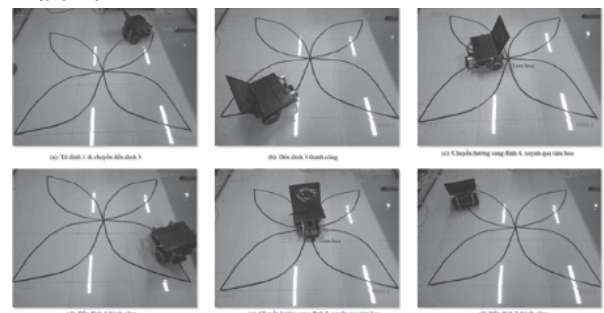
3. KẾT QUẢ VÀ THẢO LUẬN



Hình 4. So sánh thuật toán tìm đường BIT*-RL với các phương pháp khác.

Trong kịch bản mô phỏng môi trường rô bốt di chuyển chưa biết trước, các tác giả đánh giá và đối chiếu các quỹ đạo do mô hình đề xuất tạo ra với các quỹ đạo do thuật toán BIT* tạo ra. Dựa trên xác suất của các vùng trong bản đồ, thuật toán BIT*-RL được đề xuất tạo ra các quỹ đạo ngắn hơn đáng kể so với thuật toán RRT*. Trong Hình 4, các đường tạo ra dựa trên mô hình BIT*-RL có giá trị cost trung bình là 92.12. Trong khi với các thử nghiệm với RRT* có cost trung bình ở mức 103.6 và thuật toán RRT thông thường có cost 112.49. Rõ ràng phương pháp được đề xuất cho ra đường đi gần với đường đi lý tưởng nhất so với các cách tiếp cận chưa được cải tiến. Trong các trường hợp

bản đồ với vật cản dày đặc, RRT* đôi khi không thể tìm ra đường đi khả thi đến đích. Nhưng mô hình tìm kiếm được đề xuất BIT*-RL đều thành công khi nhanh chóng tìm ra đường dẫn rô bốt khả thi.



Hình 5. Rô bốt di động ba bánh tám quỹ đạo hoa bốn cánh dựa trên thuật toán dẫn đường BIT*-RL.

Để củng cố khả năng ứng dụng thực

tiền, thuật toán BIT*-RL được cài đặt trên mô hình rô bốt ba bánh thực hiện kịch bản chạy bám quỹ đạo hình cánh hoa bốn cánh như Hình 5. Với việc sử dụng mô hình học tăng cường trong việc hỗ trợ xác định hướng đi tối ưu, đã giúp loại bỏ các nhánh dẫn đường dư thừa của BIT*, đặc biệt tối ưu tài nguyên thông qua số điểm nút dẫn ít nhất có thể. Hệ thống điều khiển rô bốt đảm bảo tốc độ xử lý, sự ổn định trong quá trình bám chiến lược dẫn đường, sai số tại các đỉnh cánh hoa gần như rất nhỏ. Tổng kết lại, thuật toán BIT*-RL đảm bảo thực hiện thành công quá trình tìm đường đi đáp ứng nhiều nhiệm vụ đặt ra: Tìm đường khả thi, đường đi ngắn nhất, loại bỏ các đường dư thừa, tối ưu số nút tìm kiếm và làm trơn quỹ đạo.

4. KẾT LUẬN

Bài báo nghiên cứu trình bày một phương pháp lập kế hoạch đường đi hiệu quả và tối ưu cho WMR, kết hợp các chiến lược tránh va chạm và các kỹ thuật làm mịn để xác định tuyến đường dẫn đường hiệu quả nhất. Chiến lược dẫn đường kiểu lai kết hợp thuật toán BIT* cải tiến và tối ưu mẫu tìm kiếm sử dụng mô hình học tăng cường. Thuật toán lập kế hoạch đường đi BIT*-RL được đề xuất đã được xác nhận thành công về tính khả thi và thực hiện hiệu quả nhiều mục tiêu được thiết lập trong các thí nghiệm mô phỏng và xác nhận. ❖

Ngày nhận bài: **23/6/2025**

Ngày phản biện: **09/7/2025**

Tài liệu tham khảo:

[1]. V. T. Nguyen, D. N. Duong, and T. V. Dang, "Optimal Two-Wheeled Self-Balancing Mobile Robot Strategy of Navigation using Adaptive Fuzzy controller-based KD-SegNet". Intelligent Service Robotics, 2025, doi:

10.1007/s11370-025-00606-0.
 [2]. T. V. Dang, and N. T. Bui, "Obstacle Avoidance Strategy for Mobile Robot based on Monocular Camera". Electronics, vol. 12, no. 8, pp. 1932, 2023, doi: 10.3390/electronics12081932.
 [3]. T. V. Dang, "Optimization Hybrid Path Planning based on A-star Algorithm combining with DWA". MM Science Journal, vol. 10, pp. 7551-7555, 2024, doi: 10.17973/MMSJ.2024_10_2024077.
 [4]. L.S. Liu et al., "Path planning for smart car based on Dijkstra algorithm and dynamic window approach". Wirel. Commun. Mob. Comput., vol. 4, pp. 1-12, 2021, doi: 10.1088/10.1155/2021/8881684.
 [5]. T. V. Dang, and D. S. Nguyen, "Optimal Navigation Based on Improved A* Algorithm for Mobile Robot". Intelligent Systems and Networks, LNNS 752, pp. 574-580, 2023, doi: 10.1007/978-981-99-4725-6_68.
 [6]. Y. Sun et al., "Local Path Planning for Mobile Robots Based on Fuzzy Dynamic Window Algorithm". Sensor, vol. 23, no. 8260., 2023, doi.: 10.3390/s23198260
 [7]. T.V. Dang, D. S. Nguyen, N. T. Bui, "Hybrid Path Planning for Mobile Robot based on Improved A* Fusion Dynamic Window Approach". Proceedings of the International Conference on Intelligent Systems and Networks, LNNS 1077, pp. 82-88, 2024, doi: 10.1007/978-981-97-5504-2_10.
 [8]. A. D. Nguyen, T. D. Vu, Q. A. Vu, T. V. Dang, "Research on Modeling and Object Tracking for Robot Arm based on Deep Reinforcement Learning" MM Science Journal, vol. 6, pp.8459-8463, 2025, doi: 10.17973/MMSJ.2025_06_2025059.