2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Constructing and Navigating Connected Air Roads: A Safety-Critical Reinforcement Learning Approach for Multi-UAV Systems

Qihan Qi[1*], Haojie Xia[1*], Xinsong Yang[1], Jianquan Lu[2], *Senior Member, IEEE*, Xingxing Ju[1]

***Abstract*— This paper presents an integrated control method in air road navigation for multi-UAV systems, combining an efficient reinforcement learning (RL) controller with a control barrier function (CBF)-based filter that guarantees flight safety. First, an air road construction method based on arbitrary quadrilateral combinations is proposed, which enables flexible air road design. Second, two specific CBFs are designed: an air road CBF which keeps UAVs within designed air roads, and a collision avoidance CBF which prevents collisions between UAVs. Based on the CBF-based filter, the RL controller is allowed to be trained in a simple, single-agent environment, which reduces computational costs and enhances training efficiency. Furthermore, the RL reward is carefully designed, which considers both the stability during movement and the optimality of energy conservation. The performance, safety, and efficiency of the proposed approach are rigorously validated through comprehensive simulations and real-world experiments.**

I. Introduction

The widespread applications of Unmanned Aerial Vehicles (UAVs), such as logistics, emergency rescue, and urban surveillance [1]–[3], have resulted in increasingly crowded airspace, which makes it urgent to propose new technologies to ensure a safe UAV flight. Up to date, many effective management schemes have been proposed to ensure the safe UAV flight [4], one of which is designing air roads for UAVs.

However, existing methods on air roads have significant limitations, such as relying on potential functions to constrain UAVs within air roads [5], [6], which is easy to get stuck in local minima. Moreover, although these methods are effective in specific scenarios, it only considers simple rectangular air roads [5], which limits its ability to handle more complex environments that requires connectivity between roads. Even though [6] proposed an advanced method for connected roads, it requires complex segmentation methods to prevent local minima and thus restricts practical applications. Therefore, simpler and more effective approaches are needed for both constructing air roads and avoiding local minima.

Recently, Control Barrier Functions (CBFs) have emerged as an effective tool for ensuring safety in control systems [7]. Different from potential function methods, CBFs are highly compatible with various control algorithms, including Proportional-Integral-Derivative (PID) control, Model Predictive Control (MPC), and Reinforcement Learning (RL) control [8]–[10]. However, existing CBF methods are insufficient for air road navigation, as they lack methods for constructing connected air roads, let alone designing air road CBF for these structures. Therefore, it is a substantial challenge to propose a comprehensive solution that combines air-road construction and CBF design for air-road navigation.

While CBFs provide a framework for safety, designing the control policy remains a key challenge. Conventional controllers, such as the widely-used PID, often lack the adaptivity and optimality required for complex, high-dimensional systems. More sophisticated methods like MPC, while powerful, are constrained by their dependence on precise system models and significant computational demands, which can hinder real-time performance. In this context, RL emerges as a compelling alternative. It learns optimal controller directly from interaction data without requiring accurate system dynamics models, and its trial-and-error learning mechanism enables adaptation to parameter changes and external disturbances [11], [12]. However, existing RL methods face several challenges: achieving safety during training remains difficult, the widely adopted centralized training framework demands substantial computational resources, and learning stability is a critical concern that affects system control performance [13]–[17]. In this case, how to design an appropriate RL controller to handle above challenges? Furthermore, can the air road CBF be incorporated with the RL controller? These challenging issues inspired the research presented in the current paper.

This paper investigates a integrated controller for multi-UAV navigation in connected air roads. The proposed approach combines the safety guarantees of CBFs with the decentralized RL controller, ensuring safe UAV flight in connected air roads. The main contributions of this paper are as follows:

(1) A novel method for constructing connected air roads is proposed, which is based on the combination of quadrilateral road segments. Each road segment is an arbitrary quadrilateral, which serves as the basic unit of an air road. By connecting these segments, air roads with various shapes can be flexibly constructed.
(2) Based on construction methods, a novel air-road CBF is first proposed to ensure that UAVs remain within the connected air road, while a collision avoidance CBF is designed to maintain safe distances between UAVs. Both proposed CBFs are decentralized and easy to

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 62373262, 62303336, and 62403336. *Q. Qi and H. Xia contributed equally to this work. Corresponding authors: Xinsong Yang.
[1]Q. Qi, H. Xia, X. Yang, and X. Ju are with the College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China (e-mails: qiqihan@stu.scu.edu.cn (Q. Qi); 2023222055162@stu.scu.edu.cn (H. Xia); xinsongyang@163.com or xinsongyang@scu.edu.cn (X. Yang)); xingxju@scu.edu.cn (X. Ju).
[2]J. Lu is with the Department of Mathematics, Southeast University, Nanjing 210096, China (email: jqluma@seu.edu.cn).

implement.

(3) An integrated controller that combines RL controller and CBF-based filter is proposed for multi-UAV navigation control. A key advantage of our approach is the decoupling of tasks: the CBF-based filter manages inter-UAV collision avoidance, thereby allowing the RL controller to be trained in a simple, single-UAV setting. This strategy dramatically reduces computational load and simplifies the training process. Additionally, a carefully designed reward guides the RL controller to learn stable and optimal control actions.

For clarity, the key notations used throughout this paper are summarized in Table I. The remainder of this paper is organized as follows: Section II details the construction of air roads and the design of air road CBF and collision avoidance CBF. Section III discusses the integrated controller for collision-free navigation. Simulation and experimental results are presented in Section IV. Finally, Section V concludes the paper.

TABLE I: Notations

Notation	Description
Q^T	Transpose of vector or matrix Q
∂P	Boundary of a closed set P
Int(P)	Interior of a closed set P
$\mathbb{R}^n$	Set of n-dimensional Euclidean vector space
$\mathbb{R}^{n\times m}$	Set of $n \times m$ real matrix space
$\|\cdot\|$	Euclidean norm
0_n	n-dimensional zero vector
$0_{n\times m}$	$n \times m$ zero matrix
$\mathbf{I}_n$	$n \times n$ identity matrix

II. Air Road CBF

Consider a multi-UAV system with N UAVs. The dynamic model of UAV k, with $k \in \mathcal{N} = \{1, \ldots, N\}$ is defined as

$$\dot{x}_k = f(x_k) + g(x_k)u_k, \tag{1}$$

where $x_k = [p_k^T, v_k^T]^T$, p_k and $v_k \in \mathbb{R}^2$ are the position and velocity in the x-y axis of UAV k, respectively. The initial position and velocity are denoted as $\bar{p}_{k0}$ and $\bar{v}_{k0}$. $f(x_k) = [v_k^T, 0_2^T]^T$, $g(x_k) = [0_{2\times 2}, \mathbf{I}_2]^T$, $u_k \in \mathbb{R}^2$ is the controller.

Given the UAV dynamics in (1), the primary control objective is to design the controllers u_k, $k \in \mathcal{N}$, to ensure two critical safety properties: inter-UAV collision avoidance and invariance within a connected air-road network. Achieving both simultaneously, particularly within a decentralized framework using CBFs, presents a significant challenge. To address this, our work proposes a flexible air-road construction method coupled with a high-performance decentralized controller, which is augmented by a CBF-based safety filter. We begin by addressing the collision avoidance aspect of this framework.

For $\forall k, d \in \mathcal{N}$, $k \neq d$, the relative position and velocity between UAV k and UAV d are defined as $p_{kd} = p_k - p_d$ and $v_{kd} = v_k - v_d$, respectively. Inspired by the collision avoidance approach from [20], the following collision avoidance

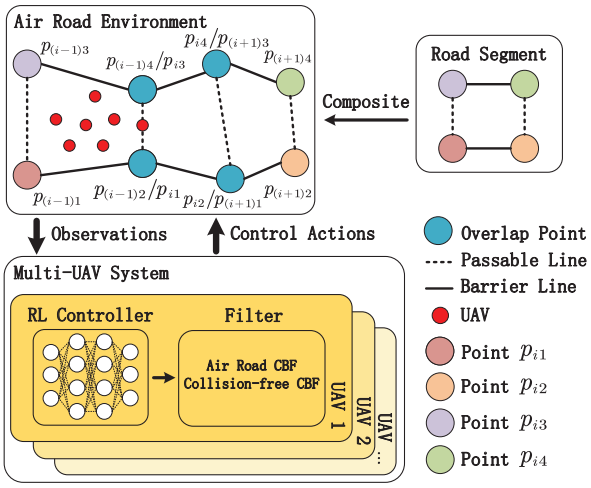


Fig. 1: The integrated control framework for multi-UAV navigation. This framework pairs a RL controller with a CBF-based safety filter. The RL controller generates efficient nominal control actions, which are then modulated by the filter to guarantee collision avoidance and bounds within the connected air roads.

constraint is designed

$$\|p_{kd}\| + \int_0^{T_{kd}} \bar{v}_{kd}(t_0 + t)dt \geq D_c, \tag{2}$$

where $\bar{v}_{kd}(t_0) = \frac{p_{kd}^T}{\|p_{kd}\|} v_{kd}(t_0)$, $v_{kd}(t_0)$ is the relative velocity at current time instant t_0, $\bar{v}_{kd}(t_0+t) = \bar{v}_{kd}(t_0) - (\delta_k+\delta_d)t$, $T_{kd} = \frac{\bar{v}_{kd}(t_0)}{\delta_k+\delta_d}$, δ_k and δ_d are the given maximum breaking acceleration constant of UAV k and d, respectively, and D_c denotes the minimum distance to keep between UAV k and d, which is a given constant.

Deriving (2), the following result is obtained

$$\|p_{kd}\| - \frac{\bar{v}_{kd}^2}{2(\delta_k + \delta_d)} - D_c \geq 0. \tag{3}$$

It should be noted that $\bar{v}_{kd}$ represents the component of the relative velocity vector along the line of sight UAV k and d. Essentially, $\bar{v}_{kd} \leq 0$ and $\bar{v}_{kd} > 0$ are the speed at which the UAVs are moving directly towards and away from each other, respectively. Therefore, the condition $\bar{v}_{kd} > 0$ is disregarded, as it indicates that the UAVs are separating. Based on this insight, the following function $b_{kd}(p_k, v_k)$ and set $\mathcal{C}_{kd}$ are determined

$$b_{kd}(p_k, v_k) = \sqrt{2(\delta_k + \delta_d)(\|p_{kd}\| - D_c)} + \frac{p_{kd}^T}{\|p_{kd}\|} v_{kd}, \tag{4}$$

$$\mathcal{C}_{kd} = \{(p_k, v_k) \in \mathbb{R}^4 | b_{kd}(p_k, v_k) \geq 0\}, \tag{5}$$

where $\|p_{kd}\| \geq D_c$ holds in practice.

The condition $(p_k, v_k) \in \mathcal{C}_{kd}$, *i.e.*, $b_{kd}(p_k, v_k) \geq 0$, signifies that if UAV k and UAV d decelerate with their respective maximum braking accelerations δ_k and δ_d, respectively, the distance between them is guaranteed to remain no less than the safety distance D_c. Therefore, by ensuring the initial state is safe, *i.e.*, $b_{kd}(\bar{p}_{k0}, \bar{v}_{k0}) \geq 0$ and designing the controller

later to render the set $\mathcal{C}_{kd}$ forward invariant, collision avoidance between UAV k and d is formally guaranteed for all time.

With the collision avoidance constraint established, we now turn to the second critical safety objective: confining the UAVs to the connected air road.

Connected air roads are constructed from road segments, which are the smallest units of the network. A road segment i is defined as a quadrilateral with four position vertices: $\tilde{p}_{i1}, \tilde{p}_{i2}, \tilde{p}_{i3}, \tilde{p}_{i4} \in \mathbb{R}^2$, where $i \in \mathcal{M}$ and $\mathcal{M} = \{1, \dots, M\}$ with M denoting the total number of segments. Each segment contains two passable paths, represented by the vectors $\tilde{p}_{i31} = \tilde{p}_{i3} - \tilde{p}_{i1}$ and $\tilde{p}_{i42} = \tilde{p}_{i4} - \tilde{p}_{i2}$, which allow UAVs to traverse. Additionally, it includes two barrier lines, represented by the vectors $\tilde{p}_{i21} = \tilde{p}_{i2} - \tilde{p}_{i1}$ and $\tilde{p}_{i43} = \tilde{p}_{i4} - \tilde{p}_{i3}$. These lines serve as the boundaries of the road segment and cannot be crossed by UAVs. The set of barrier vectors for road segment i is denoted as $R_i = \{\tilde{p}_{ij} \mid j \in \{21, 43\}\}$.

As illustrated in Fig. 1, the complete air road is constructed from a series of these quadrilateral road segments. These segments are connected at overlap points, forming locations which we denote as overlap points. By connecting these quadrilateral segments at overlap points, the air road can be configured into any desired path geometry. Additionally, it should be noted that the air road is connected seamlessly with $\tilde{p}_{i1} = \tilde{p}_{(i+1)2}$ and $\tilde{p}_{i3} = \tilde{p}_{(i+1)4}$, $\forall i, i+1 \in \mathcal{M}$.

Remark 1: The key advantage of air road construct method is its combination of simplicity and expressiveness. By connecting elementary quadrilateral units, which we call road segment, arbitrarily complex air roads can be systematically and efficiently defined.

Next, attention turns to the problem of ensuring the UAV remains within the defined air road. For $\forall i \in \mathcal{M}$, $\forall j \in \{21, 43\}$, the positive semi-definite matrix O_{ij} is defined as

$$O_{ij} = \mathbf{I}_2 - \frac{\tilde{p}_{ij}\tilde{p}_{ij}^T}{\|\tilde{p}_{ij}\|^2},$$

where $O_{ij} \in \mathbb{R}^{2\times 2}$ is an orthogonal projection matrix used in distance calculations. Based on O_{ij}, the distance vector from the UAV's position p_k to the barrier line $\tilde{p}_{ij}$ can be represented as

$$\mathbf{O}_{kij} = O_{ij}(p_k - \tilde{p}_{i(j)_2}),$$

where $(j)_2$ denotes the second digit of j, *e.g.*, if $j = 21$, then $(j)_2 = 1$.

On the basis of distance vector $\mathbf{O}_{kij}$, suppose the UAV k stays in the road segment i with a given breaking acceleration constant δ_k, and the safety distance between UAV k and the barrier line $\tilde{p}_{ij}$ is a known constant D_s, then following constraint need to be satisfied

$$\|\mathbf{O}_{kij}\| + \int_0^{T_{kij}} \bar{v}_{kij}(t_0 + t)dt \geq D_s, \tag{6}$$

where $\bar{v}_{kij}(t_0) = \frac{\mathbf{O}_{kij}^T}{\|\mathbf{O}_{kij}\|} O_{ij} v_k(t_0)$ represents the velocity of the UAV perpendicular to barrier line $\tilde{p}_{ij}$ at current time instant t_0, $v_k(t_0)$ is the current velocity of UAV k, $\bar{v}_{kij}(t_0 + t) = \bar{v}_{kij}(t_0) - \delta_k t$, $T_{kij} = \frac{\bar{v}_{kij}(t_0)}{\delta_k}$ denotes the time to make $\bar{v}_{kij}(t_0 + T_{kij}) = 0$.

Computing (6) derives that

$$2\delta_k(\|\mathbf{O}_{kij}\| - D_s) - \bar{v}_{kij}(t_0)^2 \geq 0. \tag{7}$$

For notational simplicity, let $\bar{v}_{kij}(t_0) = \bar{v}_{kij}$, $v_k(t_0) = v_k$. Intuitively, only the case that the UAV k moves toward the barrier line $\tilde{p}_{ij}$ need be considered, *i.e.*, $\bar{v}_{kij} \leq 0$. Thus, considering $\sqrt{\bar{v}_{kij}^2} = -\bar{v}_{kij} = -\frac{\mathbf{O}_{kij}^T}{\|\mathbf{O}_{kij}\|} O_{ij} v_k$, it is obtained from (7) that

$$\sqrt{2\delta_k(\|\mathbf{O}_{kij}\| - D_s)} + \frac{\mathbf{O}_{kij}^T}{\|\mathbf{O}_{kij}\|} O_{ij} v_k \geq 0,$$

where $\|\mathbf{O}_{kij}\| \geq D_s$ holds in practice.

Based on the above inequality, the air road barrier function $b_{kij}(p_k, v_k)$ and set $\mathcal{C}_{kij}$ are defined as:

$$b_{kij}(p_k, v_k) = \sqrt{2\delta_k(\|\mathbf{O}_{kij}\| - D_s)} + \frac{\mathbf{O}_{kij}^T}{\|\mathbf{O}_{kij}\|} O_{ij} v_k, \tag{8}$$

$$\mathcal{C}_{kij} = \{(p_k, v_k) \in \mathbb{R}^4 | b_{kij}(p_k, v_k) \geq 0\}. \tag{9}$$

The condition $b_{kij}(p_k, v_k) \geq 0$ provides a clear physical guarantee: it ensures that if UAV k applies its maximum braking acceleration δ_k from its current state (p_k, v_k), it is guaranteed to come to a complete stop while maintaining a distance of at least D_s from the barrier line $\tilde{p}_{ij}$. Consequently, if the initial state of the UAV is safe, *i.e.*, $b_{kij}(\bar{p}_{k0}, \bar{v}_{k0}) \geq 0$, and a controller can be designed to render the set $\mathcal{C}_{kij}$ forward invariant, the UAV will be confined to its current road segment i. Since the road segments are seamlessly connected, this confinement implies that the UAV will always remain within the connected air road.

To derive the conditions that the controller must satisfy to render both the collision-avoidance set $\mathcal{C}_{kd}$ and the road-invariance set $\mathcal{C}_{kij}$ forward invariant, we now formally define a CBF.

Definition 1: For system (1), given a set $\mathcal{C}_z$ with the continuously differentiable function $b_z(\cdot) : \mathbb{R}^n \to \mathbb{R}$, $z \in \{kd, kij\}$, the $b_z(\cdot)$ is a CBF if there exists an extended class $\mathcal{K}$ function $\alpha(\cdot)$ such that for $\forall (p_k, v_k) \in \mathcal{C}_z$, the following condition holds

$$\mathcal{L}_f b_z(p_k, v_k) + \mathcal{L}_g b_z(p_k, v_k) u \geq -\alpha(b_z(p_k, v_k)). \tag{10}$$

where an extended class $\mathcal{K}$ function $\alpha(\cdot)$ is a strictly increasing function that satisfies $\alpha(0) = 0$, the terms $\mathcal{L}_f b_z(\cdot)$ and $\mathcal{L}_g b_z(\cdot)$ are the Lie derivatives.

Let $b_{kd} = b_{kd}(p_k, v_k)$, $u_{kd} = u_k - u_d$, $b_{kij} = b_{kij}(p_k, v_k)$, selecting $\alpha(b) = \zeta b^3$, where $\zeta > 0$, ensures that $\alpha(\cdot)$ is an extended class $\mathcal{K}$ function. According to Definition 1, (4), and (8), the following two constraints must be satisfied for the controller:

1) Inter-UAV Collision Avoidance Constraint: For the collision-free CBF b_{kd} defined in (4), applying the condition (10) yields the following constraint on the relative controller

u_{kd}:

$$-p_{kd}^T u_{kd} \leq \underbrace{\frac{(\delta_k+\delta_d)v_{kd}^T p_{kd}}{\sqrt{2(\delta_k+\delta_d)(\|p_{kd}\|-D_c)}} + \|v_{kd}\|^2}_{\Delta_{kd}^1} \underbrace{-\frac{(v_{kd}^T p_{kd})^2}{\|p_{kd}\|^2} + \zeta b_{kd}^3 \|p_{kd}\|}_{\Delta_{kd}^2}. \tag{11}$$

2) Air Road Constraint: Similarly, for the air road CBF b_{kij} defined in (8), the condition (10) results in a constraint on the controller u_k for each active road barrier:

$$\underbrace{-\mathbf{O}_{kij}^T O_{ij}}_{\mathbf{A}_{kij}} u_k \leq \underbrace{\frac{\delta_k \mathbf{O}_{kij}^T}{\sqrt{2\delta_k(\|\mathbf{O}_{kij}\|-D_s)}} v_k + v_k^T O_{ij}^T O_{ij} v_k}_{\Delta_{kij}^1} \underbrace{-\frac{v_k^T O_{ij} \mathbf{O}_{kij}^T \mathbf{O}_{kij} v_k}{\|\mathbf{O}_{kij}\|^2} + \zeta b_{kij}^3 \|\mathbf{O}_{kij}\|}_{\Delta_{kij}^2}. \tag{12}$$

Remark 2: The proposed air road CBF provides a novel and practical solution for enforcing safety within complex connected air roads. Previous CBF methods [7]–[10], [20] did not provide a feasible way to propose an air road CBF for air roads with complex shapes. Existing methods are either unable to address this problem or would be impractical for arbitrarily shaped roads because they lack the modular, segment-based road construction proposed herein. One of our key innovations is a modular approach: by constructing the air road from simple quadrilateral segments, we can define a corresponding set of simple, local barrier functions. This strategy decomposes a complex global problem into a manageable set of local constraints, ensuring computational efficiency and scalability, regardless of the connected air road's complexity.

While condition (11) guarantees collision avoidance, it is a centralized constraint because the term $u_{kd} = u_k - u_d$ requires each UAV to know the other's control input in real-time, which is impractical as it requires constant communication between agents and any delay or loss of communication would make the constraint impossible to enforce. To overcome this, we seek to decentralize the constraint (11) by assigning each UAV a portion of the responsibility for maintaining safety. The original constraint can be rewritten by expanding u_{kd}:

$$-p_{kd}^T u_k - p_{dk}^T u_d \leq \Delta_{kd}^1 + \Delta_{kd}^2, \tag{13}$$

where we have used the identity $p_{kd} = -p_{dk}$. The left-hand side now represents the combined effort of both UAVs.

We decompose the combined safety effort (13) from both UAVs into individual constraints and the following effective strategy is designed to split (13) into a pair of decentralized constraints:

$$-p_{kd}^T u_k \leq \frac{\delta_k}{\delta_k+\delta_d}(\Delta_{kd}^1 + \Delta_{kd}^2), \tag{14}$$

$$-p_{dk}^T u_d \leq \frac{\delta_d}{\delta_k+\delta_d}(\Delta_{dk}^1 + \Delta_{dk}^2). \tag{15}$$

Thus, if the decentralized conditions (14) and (15) are satisfied by UAV k and UAV d, respectively, the collision-free CBF condition (11) is guaranteed.

Based on (5) and (9), for p_k in road segment i, $\forall k \in \mathcal{N}$, $\forall i \in \mathcal{M}$, the safe sets $\mathcal{C}_k^r$ and $\mathcal{C}_k^c$, which ensure that UAV k remains within the road segment i and avoids collision with other UAVs, respectively, are defined as:

$$\mathcal{C}_k^r = \bigcap_{j\in\{21,43\}} \mathcal{C}_{kij}, \quad \mathcal{C}_k^c = \bigcap_{\substack{d\in\mathcal{N}\\ d\neq k}} \mathcal{C}_{kd}.$$

The following assumptions are necessary:

- (H_1) For $\forall k \in \mathcal{N}$, the initial position-velocity pair $(\bar{p}_{k0}, \bar{v}_{k0}) \in \mathcal{C}_k^r \cap \mathcal{C}_k^c$.
- (H_2) The road segment i and h are connected seamlessly when $\tilde{p}_{i1} = \tilde{p}_{(i+1)2}$ and $\tilde{p}_{i3} = \tilde{p}_{(i+1)4}$, $\forall i, i+1 \in \mathcal{M}$.

With the CBF conditions (14) and (12), the following result is presented to keep multi-UAV system (1) within the connected air roads without collision.

Theorem 1: Suppose that assumptions (H_1)–(H_2) hold. Given a multi-UAV system with dynamics in (1), if $\forall\, k \in \mathcal{N}$, the controller u_k satisfies the CBF conditions in (12) and (14), then the multi-UAV system can stay within the air road and achieve collision avoidance.

Proof: As stated by (H_1), UAV k starts within the road segment i without collision. (H_2) guarantees that the air road is connected seamlessly through all road segments. Thus, the UAV k may remain within any road segment collision-free or approach the boundaries of the air road and/or collision.

For UAV k in road segment i, $\forall i \in \mathcal{M}$. To analyze the behavior at the boundaries, it is important to note that the boundary of $\mathcal{C}_k^r \cap \mathcal{C}_k^c$ consists of $\partial\mathcal{C}_k^r$ and $\partial\mathcal{C}_k^c$. For $(p_k, v_k) \in \partial\mathcal{C}_k^r$, it can be obtained that $(p_k, v_k) \in \partial\mathcal{C}_{kij}$ since $\partial\mathcal{C}_k^r$ consists of a specific $\partial\mathcal{C}_{kij}$, which indicates that $b_{kij} = 0$. If u_k satisfies (12), it ensures that $\dot{b}_{kij} \geq 0$ with $b_{kij} = 0$, which implies that UAV k remains within the road segment i.

For $(p_k, v_k) \in \partial\mathcal{C}_k^c$, it can be obtained that $(p_k, v_k) \in \partial\mathcal{C}_{kd}$ since $\mathcal{C}_k^c$ consists of a specific $\partial\mathcal{C}_{kd}$. If u_k satisfies (14), it ensures that $\dot{b}_{kd} \geq 0$ with $b_{kd} = 0$, which implies that UAV k avoids collisions with other UAV d. The proof is completed. ■

III. Integrated Controller

Attracted by the model-free and robustness properties of RL, this section proposes a RL controller to track the target position, and a filter that incorporates air road CBF, collision avoidance CBF, and input saturation conditions to adjust the RL control action to keep motion safe, as shown in Fig. 1. Furthermore, a navigation algorithm is designed to achieve collision-free navigation in connected air road.

For UAV $k \in \mathcal{N}$, the RL observation $s_k \in \mathcal{S}$ is defined as $s_k = [\hat{p}_k^T, \hat{v}_k^T, v_k^{rT}]^T$, where $\hat{p}_k = p_k - p_k^r$, $\hat{v}_k = v_k - v_k^r$, p_k^r denotes a target position in a navigation radius r_s, v_k^r is the target velocity. The RL controller $u_{rk} \in \mathcal{U}$ is designed as the acceleration command. Inspired by [16], the reward $r_k(t)$ is designed as:

$$\begin{aligned} r_k(t) =& - r_{k,s}(t) - r_{k,o}(t), \\ r_{k,s}(t) =& \lambda_{s1}(\|\hat{p}_k(t+1)\| - \|\hat{p}_k(t)\|) \\ &+ \lambda_{s2}(\|\tilde{v}_k(t+1)\| - \|\tilde{v}_k(t)\|), \\ r_{k,o}(t) =& \lambda_o \|v_k(t)\|, \end{aligned} \tag{16}$$

where λ_{s1}, λ_{s2}, λ_o are the given positive weights. $\|\hat{p}_k(t)\|$ and $\|\hat{v}_k(t)\|$ are a function to denote the system's energy. The stability term $r_{k,s}(t)$ guides controller to learn actions that ensure $\|\hat{p}_k(t+1)\| - \|\hat{p}_k(t)\| \leq 0$ and $\|\hat{v}_k(t+1)\| - \|\hat{v}_k(t)\| \leq 0$, which implies a decline in distance and velocity between p_k and p_k^r, v_k and v_k^r. Additionally, $r_{k,o}(t)$ is an optimality term that penalizes excessive velocities to reduce the power consumption.

Remark 3: Notably, the observation, action, and reward are based solely on individual UAV data, enabling single-UAV training that can be deployed across multiple UAVs, which significantly reduces training complexity and resource consumption compared with centralized training [14]. The decoupling of the RL tracking controller from the CBF safety module makes this architecture possible.

The action network for the RL controller u_{rk} is parameterized by a two-layer multi-layer perceptron with 128 units per layer and trained using the PPO algorithm, as detailed in [18], [19]. It should be noted that the constraints, including staying within the connected air road and avoiding collisions, are handled by CBF conditions (14) and (12). To enforce these CBF conditions, a QP-based filter is designed and the final controller u_k is denoted as $u_k = u_{rk} + u_{qk}$, where u_{qk} is the adjustment provided by filter, such that u_k satisfies CBF conditions. To guarantee that the QP is always feasible, slack variables $r_{kij} > 0$ and $r_{kd} > 0$ are introduced, inspired by [21]. The optimal adjustment u_{qk} is then obtained by solving the following QP:

$$\min_{u_{qk}, r_{kij}, r_{kd}} \quad \|u_{qk}\| + P_{kij} r_{kij}^2 + P_{kd} r_{kd}^2 \tag{17a}$$

$$\text{s.t.} \quad A_{kij}(u_{rk} + u_{qk}) \leq \Delta_{kij}^1 + \Delta_{kij}^2 + r_{kij}, \tag{17b}$$

$$-p_{kd}^T(u_{rk} + u_{qk}) \leq \frac{\delta_k}{\delta_k + \delta_d}(\Delta_{kd}^1 + \Delta_{kd}^2) + r_{kd}, \tag{17c}$$

$$\|u_{rk} + u_{qk}\| \leq \delta_k, \forall k, d \in \mathcal{N}, k \neq d. \tag{17d}$$

where P_{kij} and P_{kd} are the given positive weights. The slack variables r_{kij} and r_{kd} are introduced to ensure the QP is always feasible. The optimizer will drive them to zero if the CBF conditions can be satisfied strictly. However, when constraints conflict and would otherwise make the problem infeasible, the variables take the minimal positive values needed to relax the constraints, representing a controlled violation that guarantees a feasible control action can be found.

Additionally, the navigation radius r_s is used to limit the target position within a bounded region, which ensures bounded exploration and numerical stability during training. However, the controller cannot handle target positions outside r_s. Hence, the following multi-UAV target position navigation algorithm is introduced in Algorithm 1.

Algorithm 1 Multi-UAV Target Position Navigation

Input: Current state $x_k = [p_k^T, v_k^T]^T$, target position p_{gk}, temporary target position in radius p_k^r, target threshold ϵ, time step t, circle radius r_s, $k \in \mathcal{N}$

Output: Control action u_k

```
Set initial time count n = 0
while ||p_k − p_gk|| > ε do
    if ||p_k − p_gk|| ≤ r_s then
        Set p_k^r = p_gk
    else if ||p_k − p_k^r|| ≤ (1/2) r_s then
        Set p_k^r = p_k + ((p_gk − p_k)/||p_gk − p_k||) r_s
    end if
    Obtain observation s(nt)
    Get control action u_rk from RL controller
    Solving QP (17) and obtain u_qk
    Calculate u_k = u_rk + u_qk and execute u_k
    n = n + 1
end while
```

IV. SIMULATION AND EXPERIMENT

In this section, simulation and experiment are conducted to verify the effectiveness of the proposed methods. The following simulation is performed on a computer with the i7-13700K CPU and RTX 4060ti GPU. The experiment is executed on the Crazyflie 2.1, which is a versatile flying development platform widely used in UAVs research.

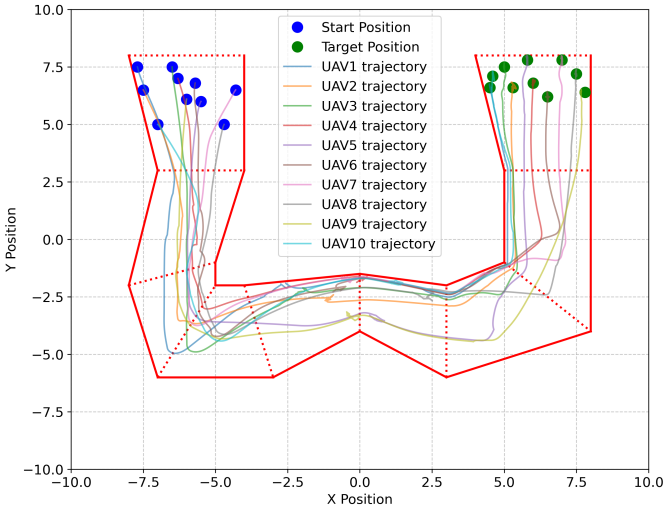


Fig. 2: Trajectories of 10 UAVs navigating through air roads with collision avoidance in simulation.

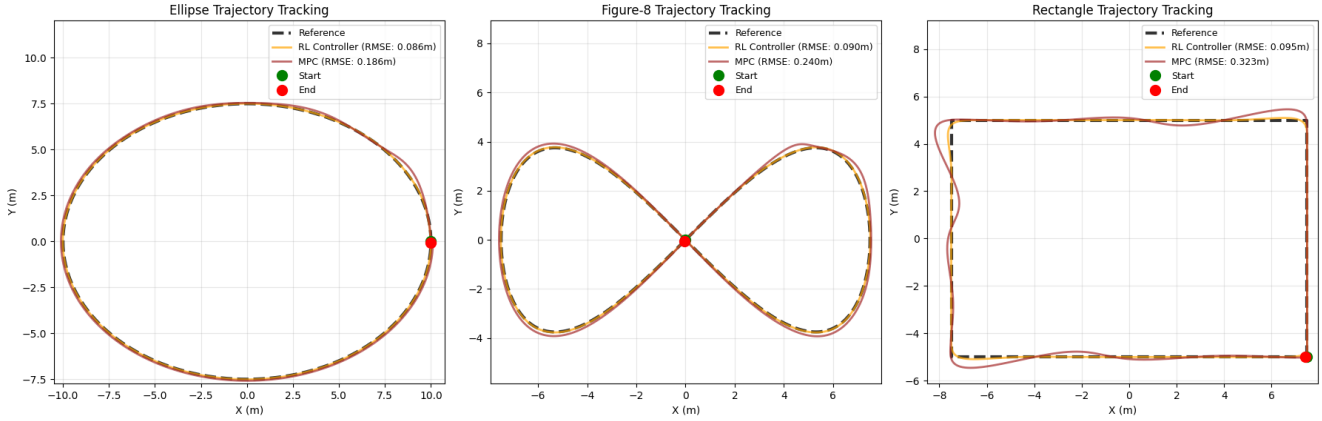


Fig. 4: Simulation comparison of RL and MPC controllers on trajectory tracking tasks for an ellipse, a figure-8, and a rectangle path.

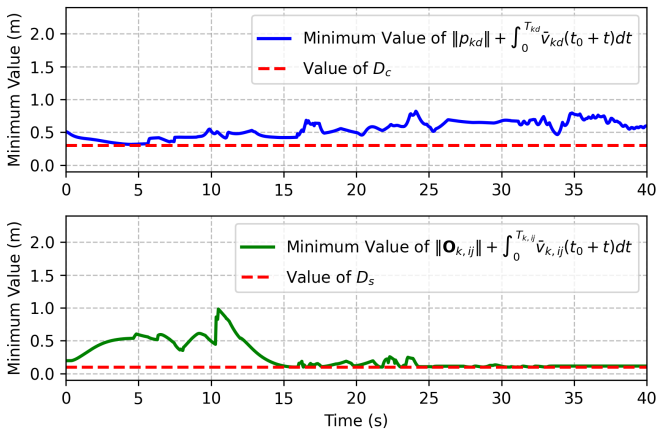
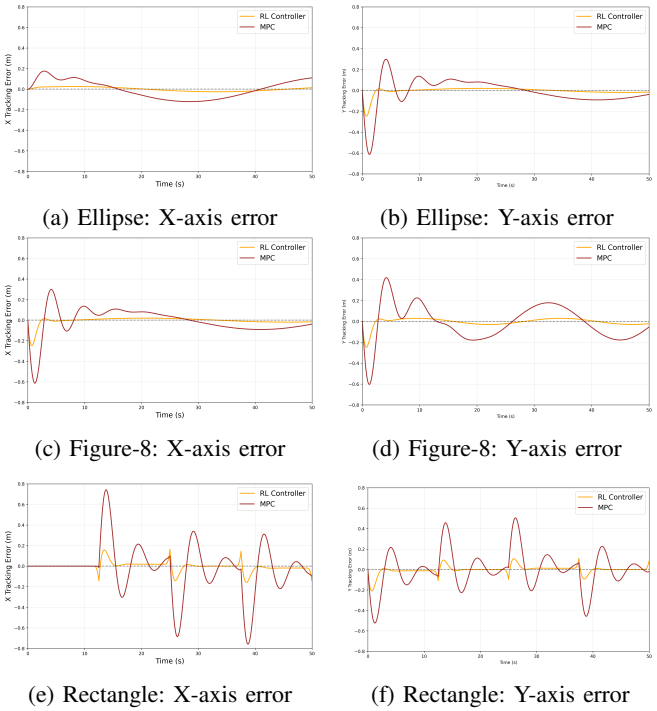


Fig. 3: Minimum values of $\|p_{kd}\| + \int_0^{T_{kd}} \bar{v}_{kd}(t_0+t)dt$ and $\|\mathbf{O}_{kij}\| + \int_0^{T_{kij}} \bar{v}_{kij}(t_0+t)dt$ compared to D_c and D_s, respectively, in simulation.



(a) Ellipse: X-axis error (b) Ellipse: Y-axis error

(c) Figure-8: X-axis error (d) Figure-8: Y-axis error

(e) Rectangle: X-axis error (f) Rectangle: Y-axis error

Fig. 5: Tracking error comparison between the RL controller and MPC controller for the ellipse, figure-8, and rectangle trajectories. Subfigures (a, c, e) show the X-axis tracking error, while (b, d, f) show the Y-axis tracking error.

A. Simulation

The simulation considers a scenario of $N = 10$ UAVs passing through air road with $M = 9$ segments. Multi-UAV system satisfies the equation (1). Algorithm 1 with RL controller and QP-based filter is applied to guide multi-UAV to the target positions. For $\forall k \in \{1, 2, \ldots, 10\}$, the acceleration $\delta_k = 5\text{m/s}^2$ and radius $r_s = 5\text{m}$. The road safety distance $D_s = 0.1\text{m}$ and collision avoidance distance $D_c = 0.3\text{m}$. The air road is shown in Fig. 2, with red solid lines depicting barrier lines and red dashed lines indicating passable routes. UAVs are initially positioned randomly along the air road, all starting with zero initial velocities. Their target positions are also specified within the air road network. The simulation step is 0.05s. Other important parameters are $\gamma = 0.95$, $\hat{\epsilon} = 0.2$, $\lambda_{s1} = 5$, $\lambda_{s2} = 0.7$, $\lambda_o = 0.1$ for PPO training, $\zeta = 0.05$ for CBFs, $\epsilon = 0.1$ for navigation algorithm.

As shown in Fig. 2, the multi-UAV system achieves the collision-free navigation from initial positions to the target positions in the connected air road using the integrated controllers. Fig. 3 shows that air road CBF and collsion avoidance CBF work effectively, since for $\forall k, d \in \mathcal{N}, i \in \mathcal{M}, j \in R_i, k \neq d$, the minimum value of $\|\mathbf{O}_{kij}\| + \int_0^{T_{kij}} \bar{v}_{kij}(t_0+t)dt$ (6) and $\|p_{kd}\| + \int_0^{T_{kd}} \bar{v}_{kd}(t_0+t)dt$ (2) are larger than D_s and D_c, respectively.

To evaluate the RL controller's performance, we conducted trajectory tracking experiments using three reference paths: an elliptical trajectory $\mathbf{p}_r(t) = [10\cos(\frac{\pi}{25}t), 7.5\sin(\frac{\pi}{25}t)]^T$, a figure-8 pattern $\mathbf{p}_r(t) = [7.5\sin(\frac{\pi}{25}t), 7.5\sin(\frac{\pi}{25}t)\cos(\frac{\pi}{25}t)]^T$, and a

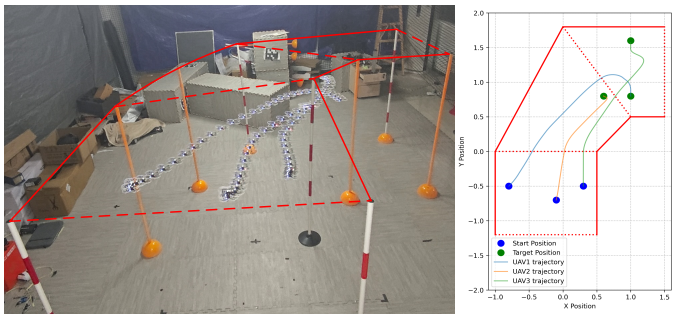


Fig. 6: Trajectories of 3 UAVs navigating through air roads with collision avoidance: left shows the real experiment, and right shows the simulation.

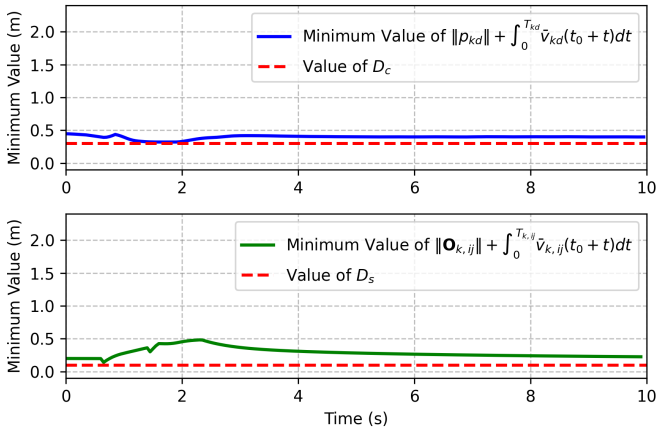


Fig. 7: Minimum values of $\|p_{kd}\| + \int_0^{T_{kd}} \bar{v}_{kd}(t_0+t)dt$ and $\|\mathbf{O}_{kij}\| + \int_0^{T_{kij}} \bar{v}_{kij}(t_0+t)dt$ compared to D_c and D_s, respectively, in experiment.

rectangular path with vertices at $p_{v1} = [7.5, -5]^T$, $p_{v2} = [7.5, 5]^T$, $p_{v3} = [-7.5, 5]^T$, and $p_{v4} = [-7.5, -5]^T$. The rectangular trajectory is defined piecewise as:

$$\mathbf{p}_r(t) = \begin{cases} p_{v1} + \frac{p_{v2}-p_{v1}}{\text{mod }(t,12.5)}, & \text{if } 0 \le t < 12.5, \\ p_{v2} + \frac{p_{v3}-p_{v2}}{\text{mod }(t,12.5)}, & \text{if } 12.5 \le t < 25, \\ p_{v3} + \frac{p_{v4}-p_{v3}}{\text{mod }(t,12.5)}, & \text{if } 25 \le t < 37.5, \\ p_{v4} + \frac{p_{v1}-p_{v4}}{\text{mod }(t,12.5)}, & \text{if } 37.5 \le t < 50, \end{cases}$$

where $\text{mod }(t, 12.5)$ represents the remainder when t is divided by 12.5. Reference velocities for all trajectories were derived through time differentiation. For all trajectory types, the UAV starts from the initial position corresponding to $t = 0$ of each reference path, i.e., $[10, 0]^T$ for ellipse, $[0, 0]^T$ for figure-8, and p_{v1} for rectangle.

The tracking performance comparison between the PPO-trained RL controller and MPC is illustrated in Fig. 4, with detailed tracking errors along X and Y axes shown in Fig. 5. Root mean square error (RMSE) values were calculated by averaging three independent trials for each trajectory. As shown in Table II, the RL controller achieved superior tracking accuracy with RMSE reductions of 53.8%, 62.5%, and 70.6% for ellipse, figure-8, and rectangular trajectories, respectively, compared to MPC. Moreover, the RL controller demonstrated remarkable computational efficiency, with solution times approximately 165 times faster than MPC across all trajectory types, making it more suitable for real-time applications.

TABLE II: Performance Comparison of MPC Controller and RL Controller

RMSE (m)	Ellipse	Figure-8	Rectangle
MPC Controller	0.186	0.24	0.323
RL Controller	**0.086**	**0.09**	**0.095**
Solution Time (ms)	Ellipse	Figure-8	Rectangle
MPC Controller	14.356	14.038	13.347
RL Controller	**0.087**	**0.086**	**0.094**

B. Experiment

Limited by venue and equipment. The real experiment is realized with $N = 3$ Crazyflie 2.1 quadcopters in the connected air road, which includes $M = 3$ road segments. The multi-UAV system states are measured by the NOKOV Motion Capture System. All UAVs start at a uniform altitude of 1m with zero initial velocities. The initial positions are given as $p_1 = [-0.8, -0.5]^T$m, $p_2 = [-0.1, -0.7]^T$m, $p_3 = [0.3, -0.5]^T$m. As shown in Fig. 6, three UAVs realize collision-free navigation in the air road under integrated controllers. It can be obtained from Fig. 7 that proposed CBFs also work well in the real experiment.

In summary, according to Fig. 2–7, the effectiveness of proposed integrated controller with navigation Algorithm 1 is verified in both simulation and experiment, which demonstrates the validity of Theorem 1.

V. Conclusions

In this paper, we develop an integrated controller for decentralized multi-UAV navigation within connected air roads, which pairs a learning-based RL controller with a CBF-based QP filter. Our framework introduces a novel method for constructing these air roads and a dedicated air road CBF to enforce safety. By leveraging the CBF filter, the RL controller can be trained in a simple single-agent paradigm, which dramatically reduces training complexity. Furthermore, a meticulously designed reward function ensures the learning of stable and optimal control policies. The efficacy and safety of the proposed approach are demonstrated through comprehensive simulations and hardware experiments.

References

[1] S. Wandelt, S. Wang, C. Zheng, and X. Sun, "Aerial: A meta review and discussion of challenges toward unmanned aerial vehicle operations in logistics, mobility, and monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 6276–6289, 2023.

[2] N. Zhao, W. Lu, M. Sheng, Y. Chen, J. Tang, F. Yu, and K. Wong, "UAV-assisted emergency networks in disasters," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 45–51, 2019.

[3] K. Dorling, J. Heinrichs, G. G. Messier, and S. Magierowski, "Vehicle routing problems for drone delivery," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 1, pp. 70–85, 2016.

[4] M. Gharibi, R. Boutaba, and S. L. Waslander, "Internet of drones," *IEEE Access*, vol. 4, pp. 1148–1162, 2016.

[5] Q. Quan, R. Fu, M. Li, D. Wei, Y. Gao, and K. Cai, "Practical distributed control for VTOL UAVs to pass a virtual tube," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 342–353, 2021.

[6] Y. Gao, C. Bai, and Q. Quan, "Distributed control for a multiagent system to pass through a connected quadrangle virtual tube," *IEEE Transactions on Control of Network Systems*, vol. 10, no. 2, pp. 693–705, 2022.

[7] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[8] M. Rauscher, M. Kimmel, and S. Hirche, "Constrained robot control using control barrier functions," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 279–285, 2016.

[9] J. Kim, J. Lee, and A. D. Ames, "Safety-critical coordination for cooperative legged locomotion via control barrier functions," *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 2368–2375, 2023.

[10] N. Takasugi, M. Kinoshita, Y. Kamikawa, R. Tsuzaki, A. Sakamoto, T. Kai, and Y. Kawanami, "Real-time perceptive motion control using control barrier functions with analytical smoothing for six-wheeled-telescopic-legged robot tachyon," *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 6802–6809, 2024.

[11] Y. Song, A. Romero, M. Müller, V. Koltun, and D. Scaramuzza, "Reaching the limit in autonomous racing: Optimal control versus reinforcement learning," *Science Robotics*, vol. 8, no. 82, pp. 1462–1474, 2023.

[12] C. Zhang, N. Rudin, D. Hoeller, and M. Hutter, "Learning agile locomotion on risky terrains," *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 864–871, 2024.

[13] C. Amato, "An introduction to centralized training for decentralized execution in cooperative multi-agent reinforcement learning," *arXiv preprint arXiv:2409.03052*, 2024.

[14] G. Wen, J. Fu, P. Dai, and J. Zhou, "DTAD: A new cooperative multi-agent reinforcement learning framework," *The Innovation*, vol. 2, no. 4, pp. 100162, 2021.

[15] C. Hu, G. Wen, S. Wang, J. Fu, and W. Yu, "Distributed multiagent reinforcement learning with action networks for dynamic economic dispatch," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 9553–9564, 2023.

[16] T. Westenbroek, F. Castaneda, A. Agrawal, S. Sastry, and K. Sreenath, "Lyapunov design for robust and efficient robotic reinforcement learning," *arXiv preprint arXiv:2208.06721*, 2022.

[17] B. S. Pavse, M. Zurek, Y. Chen, Q. Xie, and J. P. Hanna, "Learning to Stabilize Online Reinforcement Learning in Unbounded State Spaces," in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, PMLR, vol. 235, pp. 40015–40033, 2024.

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[19] Z. Hu, M. Limbu, D. Shishika, X. Xiao, and X. Wang, "Learning coordinated maneuver in adversarial environments," *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 740–745, 2024.

[20] L. Wang, A. Ames, and M. Egerstedt, "Safety barrier certificates for heterogeneous multi-robot systems," *2016 American Control Conference (ACC)*. IEEE, pp. 5213–5218, 2016.

[21] W. Xiao, C. Belta, and C. G. Cassandras, "Adaptive control barrier functions," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2267–2281, 2022.