Proceedings of the 4th International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2026.35355

A Next-Generation Hummingbird-Inspired Reinforcement Learning-Driven Flapping-Wing Robot

Zhixuan Zhou

Princeton International School of Mathematics and Science, Princeton International School, Princeton, USA
johnzhou811@gmail.com

Abstract. With the development of the low-altitude economy, unmanned aerial vehicle (UAV) systems have gradually become one of the important tools for environmental monitoring and field observation. Compared with fixed-wing aircraft and quadcopters, flapping-wing robots have significant advantages in maneuverability, environmental adaptability, and low-interference flight. However, most existing flapping-wing platforms face three major challenges: difficulty in scaling up, weak amphibious flight capability, and poor hovering stability. This project proposes a novel hummingbird-like flapping-wing robot that employs reinforcement learning control to achieve stable cross-medium flight and reliable hovering. The design utilizes biomimetic figure-eight wing motion and a simple three-degree-of-freedom mechanical structure to improve aerodynamic efficiency. The robot uses a crank-rocker transmission system, brushless motors, gearboxes, and micro servo motors to achieve wing flapping and attitude adjustment. Its fuselage is made of lightweight carbon fiber tubing, while the wings are made of waterproof and airtight fabric. A specially designed support leg structure enables the robot to float stably on water. To achieve control, we constructed a reinforcement learning framework that takes altitude, pitch angle, roll angle, and vertical velocity as input states and outputs flapping frequency, flapping amplitude, and tail angle. The reward function penalizes unstable altitude and attitude while encouraging efficient lift generation. Test results show that the reinforcement learning model converges well during training. The robot can accurately track the target trajectory in three-dimensional position and attitude, and stably complete takeoff, landing, and stable floating on water. This work demonstrates that biomimetic design combined with reinforcement learning can effectively improve the performance of flapping-wing robots and support cross-medium practical applications, providing a reference for future environmental monitoring and field observation.

Keywords: Flapping-wing robot, Hummingbird-inspired, Reinforcement learning, Water–air cross-medium, Bionic design

1. Introduction

Unmanned aerial vehicles (UAVs) have been widely used in environmental protection, ecological surveys, and disaster monitoring. Currently, the most common UAV platforms include fixed-wing

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.35355

UAVs, quadcopter UAVs, and ornithopter UAVs, as shown in Figure 1, from left to right.



Figure 1. Fixed-wing UAV, quadcopter and ornithopter existing on the market

Each type of UAV has its advantages and disadvantages. Fixed-wing UAVs rely on fixed aerodynamic principles, resulting in high cruising efficiency, but their flight mode flexibility is limited, their maneuverability is poor, and they require runways for takeoff and landing. Quadcopter UAVs, based on the vertical lift principle of multi-rotors, can achieve stable hovering and good accuracy, but they are susceptible to environmental interference and cannot operate on water.

In contrast, ornithopter-like bird-like aircraft employ a bio-inspired flapping mechanism, resulting in superior overall performance: they support multiple flight modes, including cruising, gliding, and hovering; they have excellent environmental adaptability, low noise and interference; they possess high maneuverability and wind energy adaptability for close-range operations; and uniquely, they have the ability to take off and land on water, making them the most flexible choice for multi-scenario missions [1, 2].

Despite these advantages, current ornithopter-like robots still have some significant limitations. First, many designs are difficult to scale and lack sufficient sensors to meet the demands of practical tasks. Second, few robots are capable of cross-medium flight, and adaptive control for cross-medium flight is lacking, making smooth transitions from water to air difficult [2-4].

To address these issues, this paper proposes a novel, reinforcement learning-based, hummingbird-inspired flapping-wing robot. The goal is to create a scalable platform capable of stable flight and supporting aquatic operations [1, 2, 5, 6].

This work contains two major innovations:

• A biomimetic mechanical structure employing a figure-eight airfoil trajectory [3, 7] and a controllable three-degree-of-freedom system.

• A reinforcement learning control method is proposed that can handle unsteady aerodynamics without relying on complex mathematical models.

Finally, experiments demonstrate that the two innovations mentioned above can effectively improve the system's adaptability and robustness.

2. Flapping-wing robot structure

2.1. Overall structure

The flapping-wing robot's design mimics the figure-eight wing trajectory of a hummingbird's natural flight pattern. Employing a simplified mechanical structure, it possesses three controllable degrees of freedom, enabling independent adjustment of wing flapping, pitch, and roll. This simplified mechanical structure improves reliability, reduces weight, and makes the system easier to manufacture, control, and maintain, while still supporting stable flight in the air and hovering over water [3].

As shown in Figure 2, the flapping-wing robot consists of three main functional components: the wing assembly, the fuselage with a power transmission system, and the tail control unit. The wing is

DOI: 10.54254/2753-8818/2026.35355

constructed with a lightweight frame structure and waterproof, airtight wing surfaces. This lightweight and waterproof design allows the flapping wing to generate stable lift during the transition from water to air. The flapping wing's power system consists of A2212 brushless motors, driven by a gear set and a crank-rocker mechanism. The cooperation of connecting rods and spherical sliding bearings achieves smooth, multi-degree-of-freedom wing movement, mimicking the figure-eight trajectory of a hummingbird in natural flight. Two MG90S micro servo motors control the tail wing. One servo motor adjusts the pitch angle via an auxiliary crank-rocker mechanism, while the other controls the roll angle. This configuration gives the robot three controllable degrees of freedom, ensuring efficient flapping motion and precise attitude control, thereby achieving stable flight and cross-medium operation.

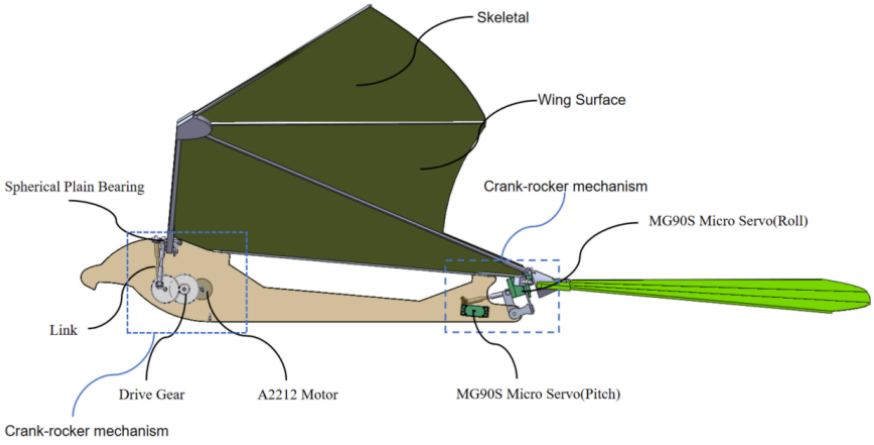

Figure 2. General drawing of flapping-wing robot

2.2. Flapping-wing profile structural framework

Figure 3 details the structural framework of the flapping-wing robot. Figure 3(a) shows the physical prototype after the design was completed, featuring a symmetrical biplane layout with a wingspan of 1750 mm and a vertical height of 950 mm. Figure 3(b) is an isometric view of the overall assembly using SolidWorks, revealing the streamlined design of the wings.



Figure 3. The structural framework of the flapping-wing robot (a) Physical prototype(b) 3D design of the overall assembly

Figure 4 is a close-up of the internal structure of the flapping-wing robot, including the crank-rocker mechanism, drive gear assembly, and servo-controlled tail fin component. The core of the flapping-wing robot's modular design lies in its lightweight frame, symmetrical wing layout, and compact drive system. These three modules constitute the robot's body. Through a balanced structural design, it supports efficient flapping motion and precise attitude control, meeting the fundamental requirements for stable hovering and cross-medium operation.

DOI: 10.54254/2753-8818/2026.35355

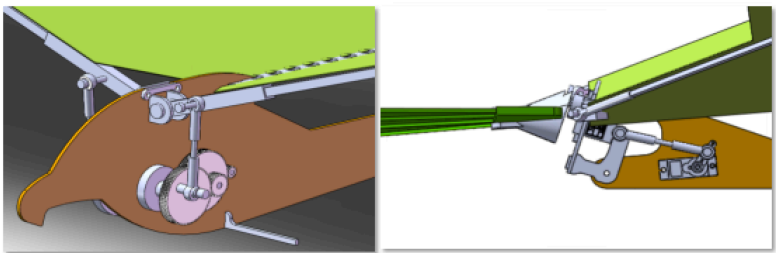

Figure 4. Close-up of the internal structure of the flapping-wing robot

3. Mechanical design and performance calculation

3.1. Wing drive section

Figure 5 shows the crank-rocker four-bar linkage used to drive the flapping wing of the robot, used to verify the motion trajectory and mechanical characteristics of the wing flapping motion.

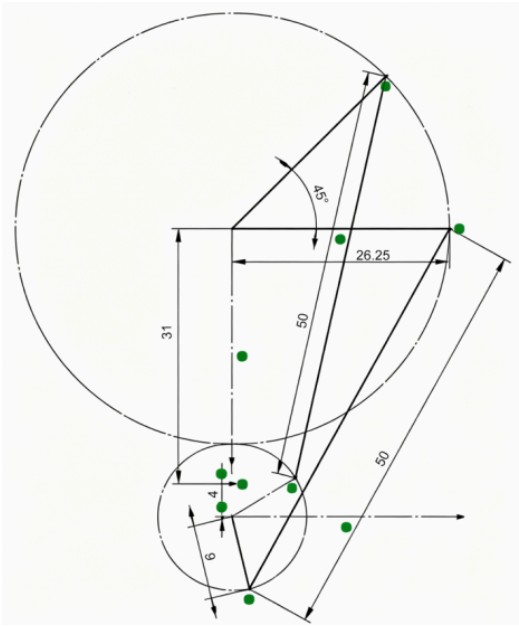

Figure 5. Driving crank-rocker mechanism

The wing drive system is designed to generate sufficient lift and torque for stable flapping motion. Key kinematic and aerodynamic parameters are calculated as follows. First, calculate the tip wing speed,which is determined by the flapping frequency and wing radius, the formula is as follow and the calculation yields a result of 6.6 meters per second.

$$v_{tip}=2\pi fr \tag{1}$$

The flapping-wing lift is then computed using the aerodynamic lift formula the calculation result is 11.23 Newtons.

$$L=0.5*C_L\rho v_{tip}^2 S \tag{2}$$

DOI: 10.54254/2753-8818/2026.35355

It is derived from the following formula the corresponding wing aerodynamic moment is 9.83 Newtons

$$T_1 = L * r \tag{3}$$

The drive mechanism uses a crank-rocker configuration with the following dimensions shown in Table 1

Table 1. Crank-rocker structure dimensions

Item	Parameters
Crank radius	9 mm
Connecting rod length	50 mm
Rocker length:	26.25 mm
Frame distance:	31 mm
Crank angle	45°

According to the principle of kinematics of mechanisms, the instantaneous torque amplification ratio and the required motor torque can be derived through formulas 4 and 5. The results are 0.276 and 35.6 Nm respectively.

$$i_M = \frac{T_1}{T_M} = \frac{R * \sin(\alpha+\delta)}{b * \sin\beta} \tag{4}$$

$$T_M = \frac{T_1}{i_M} \tag{5}$$

The selected brushless motor is shown in Figure 6 (a). Its key parameters are rated torque of 0.1 N·m, maximum torque of 0.3 N·m, rated voltage of 9 V, and rated speed of 10,000 r/min. These values define the rated operating conditions and performance limits of the motor to meet the working requirements of the flapping-wing robot.

The selected reducer is shown in Figure 6 (b), with core parameters of a reduction ratio of 20:1, a peak output speed of 1000 rpm, a rated torque of 3 Nm, and a peak torque of 7 Nm. To verify the selection, the actuation frequency was verified based on crank rotations of 45° and 90° per cycle, with 4 cycles per revolution. The input speed was calculated by multiplying the motor speed of 10000 rpm (divided by 15.11, then by the reduction ratio of 20) by 4, resulting in a frequency of 2 Hz, satisfying the 1 Hz design requirement. Furthermore, the torque check assumed a motor torque of 0.15 Nm. The calculated output torque was 45.33 Nm, greater than the required 35.6 Nm, thus a safety factor of 1.27. These calculations confirm that the gearbox meets the design requirements.



Figure 6. Selection of drive mechanism for crank-rocker mechanism (a) Drive motor A2212 (b) Reduction gearbox

DOI: 10.54254/2753-8818/2026.35355

3.2. Tail direction control section

The core control unit of the tail section adjusts the robot's pitch and roll attitude, providing the aerodynamic torque required for stable flight. As shown in Figure 7, the four-bar crank-rocker mechanism driven by the tail fin converts the rotational motion of the servo motor into the reciprocating oscillation of the tail fin.

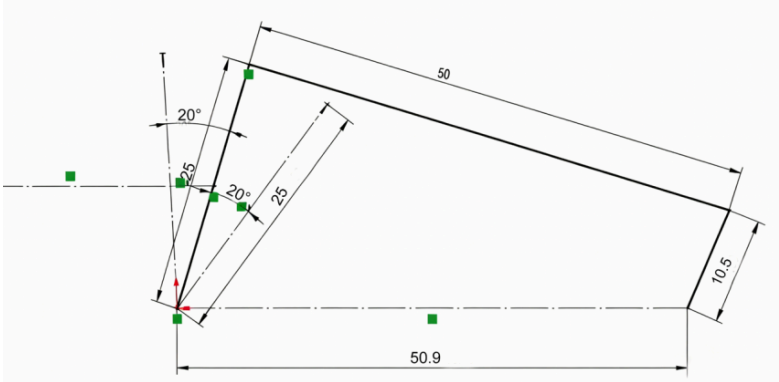

Figure 7. Tail wing driving crank-rocker mechanism

The tail fin operates at a lower effective speed relative to the main wingtip; its speed is calculated using the following formula, yielding a speed of 3.3 m/s

$$v_{tail}=0.5v_{tip} \tag{6}$$

Based on this speed, the lift generated by the tail fin can be calculated using the aerodynamic lift formula, resulting in 0.45N.

$$L=0.5*C_L\rho v_{tail}^2 2S \tag{7}$$

The corresponding aerodynamic torque acting on the tail fin is determined to be 0.09Nm, calculated using the following formula

$$i_M=\frac{T_1}{T_M}=\frac{R*\sin(\alpha+\delta)}{b*\sin\beta} \tag{8}$$

$$T_1=\frac{2}{3}L*l \tag{9}$$

To convert servo rotation into tail wing motion, the tail wing drive mechanism also employs a crank-rocker mechanism with the following geometric parameters: crank radius 10.5mm, connecting rod length 50mm, rocker length 25mm, frame spacing 50.9mm, and a typical crank angle of 40°. Using these parameters, the instantaneous torque amplification ratio of the mechanism is derived to be 0.69, as shown in the formula.

$$T_M=\frac{T_1}{i_M} \tag{10}$$

The actuator selected for tail control is the TowerPro MG90S micro servo motor shown in the figure 8, with a rated torque of 0.16 N·m, an operating voltage range of 4.8–6.0 V, and an angular velocity of 600°/s. Comparing the rated torque of the servo motor with the calculated required

DOI: 10.54254/2753-8818/2026.35355

torque, we find that T = 0.16 Nm > 0.13 Nm, corresponding to a safety factor of 1.23. This confirms that the MG90S servo motor provides sufficient torque for pitch and roll control applications.



Figure 8. Drive motor MG90S (pitch&roll)

3.3. Frame, materials and overall performance

In summary, the final design of the flapping-wing robot utilizes a carbon fiber tubular frame, characterized by high strength and lightweight. The wing surface is covered with airtight fabric, ensuring effective lift generation while also providing waterproofing, enabling operation in environments ranging from water to air. Its dynamic performance meets the requirements for flight and water-based hovering scenarios; performance parameters are shown in Table 2.

Table 2. Key performance parameters of the flapping-wing robot

Parameter	Specification
Total weight	1.0 kg
Wingspan	1750 mm
Flapping frequency	1–2 Hz
Wingtip velocity	5.5 m/s
Flight velocity	7 m/s
Up-flapping angle	0–45°
Tail pitch deflection	±20°
Tail roll deflection	±15°
Actuators	A2212 motor, MG90S servos

4. Water surface hovering design

To achieve stable cross-medium operation, the flapping-wing robot is equipped with a support leg structure specifically designed for hovering on water. This design allows the robot to safely take off, land, and stop on water while maintaining stability at the water-air interface.

4.1. Analysis of water surface operation

A mechanical analysis was conducted on the flapping-wing robot's floating state on water. The figure 9 shows a mechanical model of the support leg at the water-air interface. The forces acting on the support leg when floating on water include hydrostatic pressure, surface tension, and viscous

DOI: 10.54254/2753-8818/2026.35355

drag. The resultant force of these forces must balance the robot's lift and buoyancy to achieve stable hovering. For a flapping-wing robot with a mass of 1 kg, a wingspan of 1.75 m, and a flapping frequency of 1 Hz, equipped with two support legs with a diameter of 50 mm, the resultant force of hydrostatic pressure, surface tension, and air drag is approximately 0.18 N.

Simultaneously, the buoyancy generated by the support leg is calculated to be 0.1926 N. During hovering, the vertical force balance requires that the robot's lift plus buoyancy equals its total weight plus the resultant force at the water-air interface. This results in a minimum lift of approximately 9.8 N required to maintain stable hovering. The instantaneous lift generated during flapping wing descent is approximately 11.23 N, exceeding the minimum lift requirement, thus enabling the robot to achieve stable hovering on the water surface.

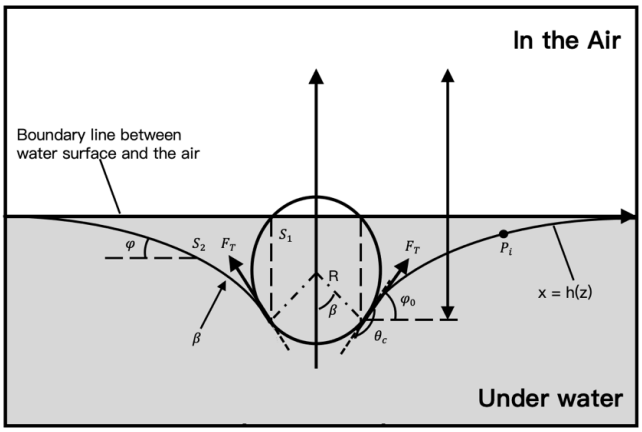

Figure 9. Forces on a circular-section leg floating on water surface and deformation of water-air interface

4.2. Supporting leg design

As shown in Figure 10, the support leg structure is designed to balance lightweight design with sufficient buoyancy. The use of circular cross-section leg dimensions provides sufficient surface area to generate surface tension and buoyancy while maintaining a compact structure to minimize flight drag. This structure is optimized to ensure the robot remains level and stable while floating, preventing capsizing or sinking. Furthermore, the design considers the deformation of the water-air interface and the force distribution on each leg, ensuring reliable operation of the robot during both static floating and dynamic takeoff and landing operations.

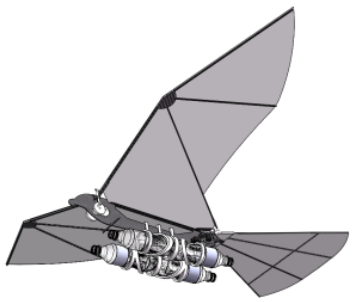

Figure 10. Design on supporting leg

DOI: 10.54254/2753-8818/2026.35355

5. Reinforcement learning based control system

5.1. Overall RL framework design

To achieve cooperative control optimization for flapping-wing robots, this paper proposes a solution framework based on reinforcement learning (RL). Its overall logic operates within a Markov Decision Process (MDP) framework, specifically consisting of three parts: environment, encoder, and optimization algorithm (see Figure 11).

The framework's operational mechanism lies in the deep coupling between the agent and the physical environment. The agent directly intervenes in the environment by outputting control actions (such as adjusting frequency or deflection angle), while the environment returns a reward value based on a pre-defined physical model. This value comprehensively considers the robot's hovering accuracy, flight stability, and energy consumption.

The environment module is primarily responsible for formally describing physical constraints and task indicators. It not only encompasses complex aerodynamic models but also associates the robot's current physical configuration with control decisions through state-space mapping.

To address the high-dimensional characteristics of sensor data, this study designs a dedicated encoding layer. Through a multi-layered neural architecture (such as convolution or graph connections), the original wing position and environmental observations are compressed into compact feature representations. This processing method effectively filters redundant noise and extracts more relevant information beneficial for policy learning.

Finally, a reinforcement learning algorithm executes the policy iteration. This process aims to maximize expected return, enabling the agent to extract the optimal control law from experience through sequential decision learning. This end-to-end learning paradigm significantly improves the operational reliability of flapping-wing robots in complex dynamic environments.

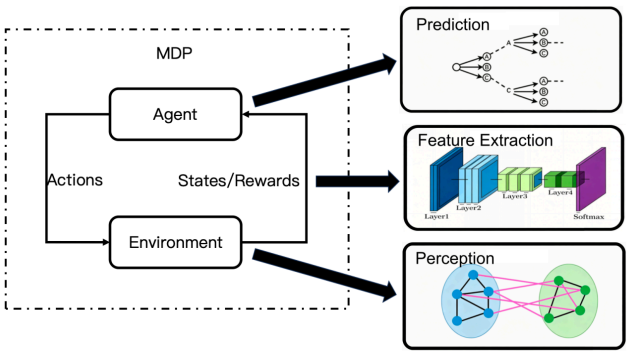

Figure 11. Reinforcement learning framework for combinatorial optimization

5.2. PPO control framework

As shown in the figure 12, the Proximal Policy Optimization(PPO) control algorithm aims to achieve stable adaptive control of a flapping-wing robot by combining the Actor-Critic framework and a dedicated reward-penalty calculation system.

The process first inputs environmental observation data into the Actor model, which acts as a policy network to generate control actions At, such as flapping frequency, amplitude, and tail wing deflection angle. These actions are then passed to the Critic model, which estimates the state value

DOI: 10.54254/2753-8818/2026.35355

Vt, providing a stable baseline for evaluating the long-term performance of the current policy and thus reducing training variance.

Simultaneously, the reward-penalty calculation system, comprising a reward model and a reference model, generates a reward signal Rt based on predefined performance metrics, such as flight stability and lift efficiency. The reference model prevents the Actor from deviating too far from the original policy distribution, ensuring training stability.

Finally, using the generated actions At, the estimated value Vt, and the reward signal Rt, the PPO loss function is calculated and backpropagated to update the parameters of the Actor model, iteratively improving the control policy to balance performance, stability, and consistency with robot operation requirements.

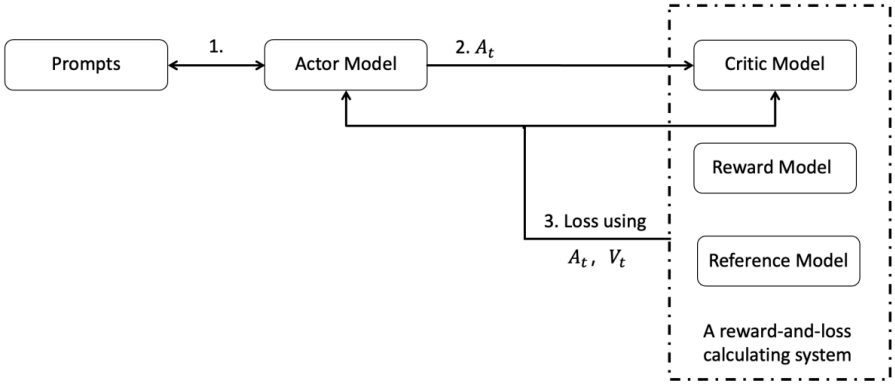

Figure 12. PPO training pipeline for the flapping-wing robot control system

5.3. Training architecture and process

Figure 13 shows the simulation environment of the reinforcement learning control system. The screenshot illustrates the evolution of the robot's flight behavior in a physics-based environment during different training phases (including takeoff, attitude adjustment, and stable hovering).

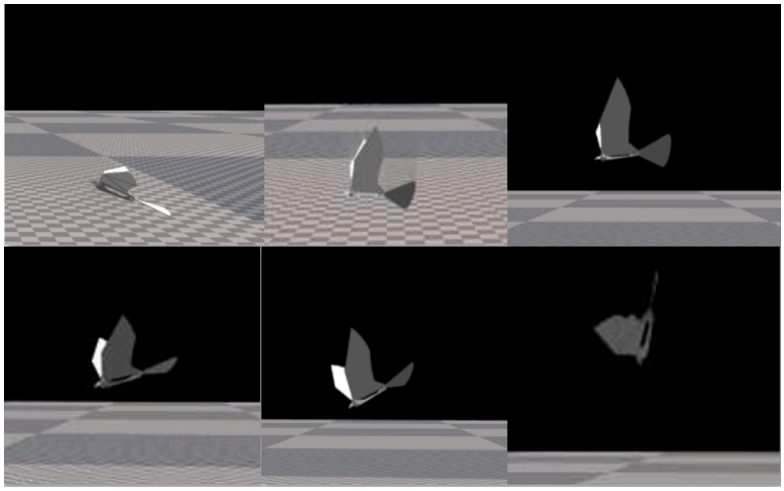

Figure 13. Screenshot of the simulation environment of (RL) training process

During reinforcement learning, the core of the agent includes its input state space, containing real-time flight parameters such as altitude, pitch angle, roll angle, and vertical velocity, comprehensively representing the robot's dynamic state. Its output action space consists of

DOI: 10.54254/2753-8818/2026.35355

continuous control commands used to control flapping frequency, flapping amplitude, and tail fin deflection angle, directly adjusting the robot's aerodynamic performance and attitude.

Simultaneously, a well-designed reward function balances penalties for altitude deviation, attitude fluctuations, and surface drag with the incentive of effective lift, guiding the agent to learn stable, efficient, and adaptive control strategies through repeated interaction with the simulation environment.

6. Experimental results and analysis

6.1. Reinforcement learning training performance

Figure 14 shows the training curves of this reinforcement learning process, illustrating key metrics including Actor loss, Critic loss, and average stage reward. As can be seen from the figure, the Actor loss exhibits a clear decreasing trend, with variance gradually decreasing during training steps, indicating that the policy network continuously improves its action selection strategy and aligns it with the reward objective.

The Critic loss shows a sharp upward trend in the early stages of training, a typical phenomenon of the value function rapidly adapting to policy changes, followed by a gradual decrease and stabilization, reflecting that the Critic network is converging to an accurate value estimate.

Simultaneously, the average reward curve shows a stable upward trend, gradually increasing from a low initial value to a stable plateau value, indicating that the agent is continuously learning effective control strategies, and the overall performance of the flapping-wing robot is constantly improving. The complementary trends of these three metrics collectively verify the stability of the training process, the convergence of the algorithm, and the effectiveness of the learned strategy for the target control task.

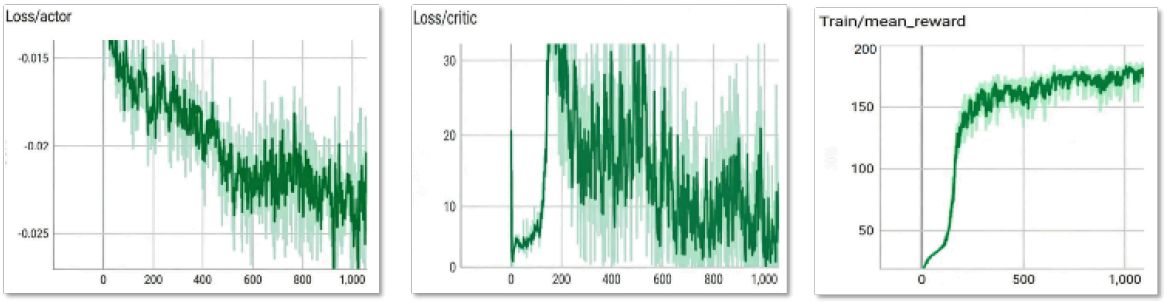

Figure 14. Training curve of reinforcement learning

6.2. Trajectory and attitude tracking performance

Figure 15 compares the tracking results of the robot's actual flight trajectory with the predefined reference trajectory in both position and attitude dimensions, demonstrating the trajectory tracking performance of the reinforcement learning-based controller. As can be seen from the X, Y, and Z axes in the figure, the actual flight path closely matches the designed trajectory throughout the entire 100-second task, with only minor deviations. Similarly, the attitude tracking results (including roll, pitch, and yaw angles) also show that the robot's actual attitude matches the reference commands well, without significant drift or oscillation.

These results confirm that the reinforcement learning-based control system achieves high-precision tracking in both position and attitude control, effectively enabling the flapping-wing robot

DOI: 10.54254/2753-8818/2026.35355

to stably and accurately track the predetermined flight path.

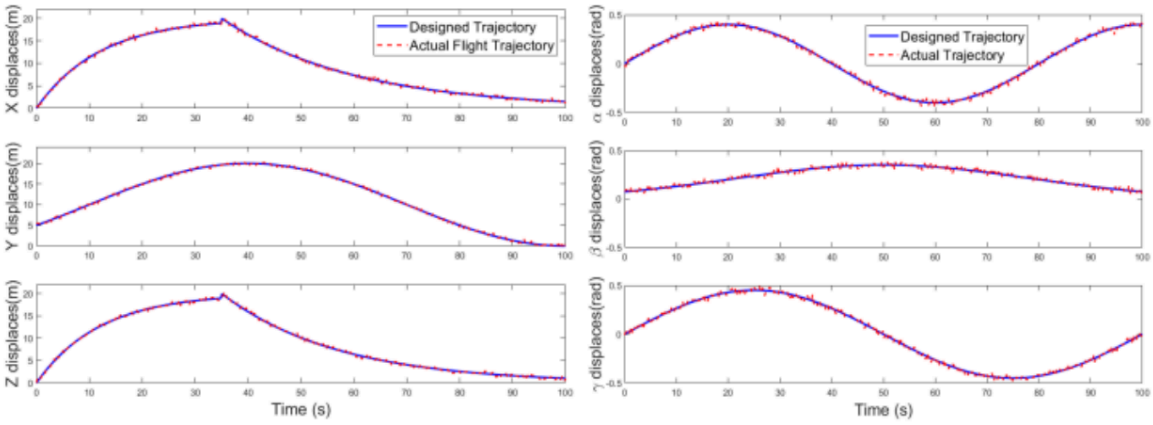

Figure 15. Control performance after RL model (a) Attitude diagram of flapping-wing (b) Flapping-wing position diagram

6.3. Real-world flight and water surface tests

Building on the simulation results, outdoor field tests were conducted to further validate the controller's performance under real-world conditions.

Figure 16 shows the aerial flight test, evaluating the robot's complete operational cycle from takeoff to landing. The test procedure began with manual release of the robot to provide initial velocity. After release, the reinforcement learning controller activated, adjusting the flapping frequency and tail wing deflection angle to initiate autonomous flight. Subsequently, the robot entered a stable hovering state, and finally, a remote controller guided the robot to a controlled landing, gradually decreasing altitude while maintaining attitude stability. Throughout the process, the robot successfully completed all stages of the task without significant oscillations or instability, confirming that the learned strategy can be effectively transferred from the simulation environment to the actual flight environment.



Figure 16. Illustration of aerial flight of flapping-wing robots

In addition to the aerial flight test, the robot's water-floating capability was also tested, as shown in Figure 17. The robot is equipped with a lightweight buoyancy module, enabling it to maintain a stable floating state on calm water, keeping the fuselage level and the wings not in contact with the

DOI: 10.54254/2753-8818/2026.35355

water. This configuration allows the robot to land and remain on the water, thus performing multi-stage tasks combining aerial flight and water operations. The floating test validated the robot's mechanical design and weight distribution, confirming that it can safely remain on the water surface without sinking or capsizing, thus expanding its operational scenarios beyond pure flight.

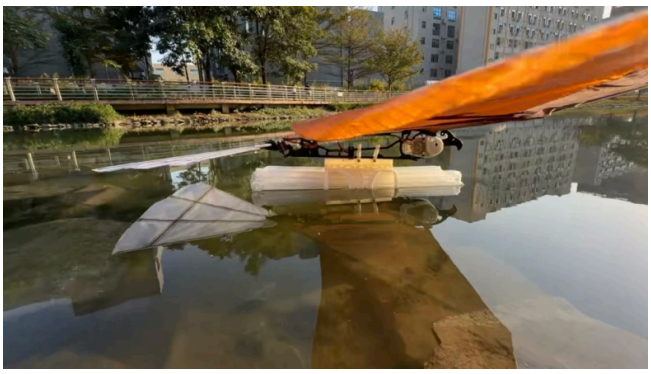

Figure 17. Illustration of flapping-wing robot floating on the water surface

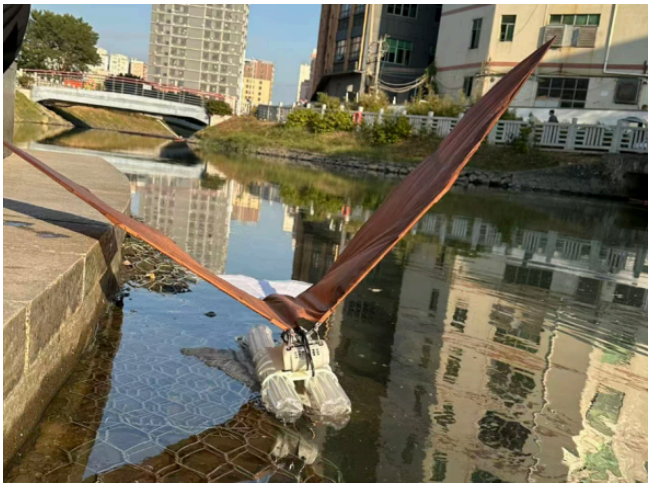

Figure 17. Illustration of flapping-wing robot floating on the water surface

7. Conclusion and future work

7.1. Conclusion

The experimental results of this work ultimately proved the two core innovations proposed: a biomimetic flapping wing mechanism and a reinforcement learning-based adaptive control system.

First, designed and manufactured a 1.75 meter wingspan, 1-kilogram flapping-wing aircraft. At a flapping frequency of 1 Hz, the prototype generated over 11.23 Newtons of lift, sufficient to support its own weight and carry a small amount of additional payload. More importantly, the aircraft can hover over water and switch between water and air, a feature uncommon in flapping-wing aircraft of similar size.

To achieve stable autonomous flight and hovering, a reinforcement learning control method was employed. Unlike traditional model-based controllers, reinforcement learning does not rely on precise aerodynamic models; instead, it learns interactively, proving well-suited to the highly unstable dynamics of flapping-wing aircraft.Experimental results show that the aircraft maintains

DOI: 10.54254/2753-8818/2026.35355

good position and attitude tracking accuracy. This performance level meets the basic requirements for low-speed, near-ground operations—such as environmental monitoring or inspection tasks near water. The findings provide practical reference for future research on autonomous hovering and more precise control of flapping-wing aircraft, particularly those aiming to achieve cross-medium capabilities [7, 8].

7.2. Future work

While the current research offers some insights, it is only a first step. Looking ahead, several directions warrant exploration. First, the current structure and flapping wing mechanism can be further improved. Although the prototype has generated sufficient lift, there is still room for improvement in lift efficiency and endurance, especially in cross-medium flight where energy consumption is often higher. Small modifications to the transmission system or wing materials could lead to significant performance improvements [6, 7].

Regarding control, the reinforcement learning algorithm used in this paper, while feature-rich, is far from optimal. Future research could focus on improving its stability and resistance to interference in real-world environments, such as gusts of wind or ocean waves. This involves more than just adjusting hyperparameters; it may require redesigning the reward mechanism or introducing domain randomization during training. Greater robustness will make the aircraft more suitable for outdoor operations [4, 5].

References

[1] Warrick, D. R., Tobalske, B. W., & Powers, D. R. (2005). Aerodynamics of the hovering hummingbird. Nature, 435(7045), 1094-1097.
[2] Chen, Yufeng, et al. "A biologically inspired, flapping-wing, hybrid aerial-aquatic microrobot." Science robotics 2.11 (2017): eaao5619.
[3] He J, Zhang Y, Feng J, et al. A novel aerial-aquatic unmanned vehicle using flapping wings for underwater propulsion [J]. Biomimetics, 2024, 9(10): 581.
[4] Meng R., Song B., Xuan J., et al. Design and Verification of a Large-Scaled Flapping-Wing Aircraft Named "Cloud Owl" [J]. Applied Sciences, 2023, 13(9): 5667
[5] Wu C, Xiao Y, Zhao J, et al. JingWei: A Waterfowl-inspired Flapping-wing Robot with Multimodal Aerial-aquatic Mobility [J]. IEEE Robotics and Automation Letters, 2025.
[6] Tang D., Huang X., Che J., et al. Quantitative analysis of the morphing wing mechanism of raptors: Analysis methods, folding motions, and bionic design of Falco Peregrinus [J]. Fundamental Research, 2024, 4(2): 344-352
[7] Wang C., Liu Y., Xu D., et al. Aerodynamic performance of a bio-inspired flapping wing with local sweep morphing [J]. Physics of Fluids (1994), 2022, 34(5): 1-8
[8] Liu K., Song B., Chen A., et al. Effects of dynamical spanwise retraction and stretch on flapping wing forward flights [J]. Chinese Journal of Aeronautics, 2024, 37(4): 181-202