2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Variable Admittance Interaction Control of UAVs via Deep Reinforcement Learning

Yuting Feng, Chuanbeibei Shi, Jianrui Du, Yushu Yu*, Fuchun Sun, Yixu Song

Abstract— **A compliant control model based on reinforcement learning (RL) is proposed to allow robots to interact with the environment more effectively and autonomously execute force control tasks. The admittance model learns an optimal adjustment policy for interactions with the external environment using RL algorithms. The model combines energy consumption and trajectory tracking of the agent state using a cost function. Therein, an Unmanned Aerial Vehicle (UAV) can operate stably in unknown environments where interaction forces exist. Furthermore, the model ensures that the interaction process is safe, comfortable, and flexible while protecting the external structures of the UAV from damage. To evaluate the model performance, we verified the approach in a simulation environment using a UAV in three external force scenes. We also tested the model across different UAV platforms and various low-level control parameters, and the proposed approach provided the best results.**

I. INTRODUCTION

UAVs have become increasingly popular because of their high efficiency and sensitivity coupled with low costs. At the same time, UAVs can be exposed to dangerous or toxic environments, as they can complete interactive tasks in nearly every conceivable workspace. Interaction tasks may be diverse and in unknown environments. As such, UAVs are used to manipulate unstructured environments through contact, which may include assembly tasks [1], peg-in-hole tasks [2], human-robot co-manipulation tasks [3], assisting responders in search and rescue scenarios [4], impacting with a vertical surface [5], connecting multiple UAVs with load [6], etc. The dynamics of the aerial manipulator with unknown disturbances was analyzed [7] , which provided much probability for contacting with an environment. Conducting complex tasks brings challenges for UAVs, especially if the environment is unstructured and changeable, which requires advanced interaction control. The force control characteristics of UAVs can be described by the inertia, stiffness, and damping parameters. To obtain good control performance, it is necessary to have a deep understanding of the controller design and manually adjust the parameters based on the task characteristics. Cartesian impedance control is a classical interaction approach that was used by Lippiello et al. [8]. This provides a dynamic relationship between the external generalized forces acting on the structure and the system motion.

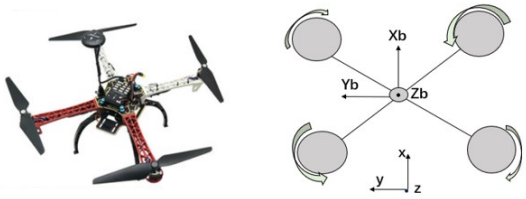

Fig. 1. A UAV in the X configuration with four motors and four rotors. The direction of motor rotation and the axis are shown in the picture.

Interaction control can be used in Cartesian space to control interactions of the end-effector with the environment [9]–[11], such as with haptic exploration [12]. This can also be used in joint spaces to enhance safety [13]–[15]. Cartesian interaction control with null-space stiffness is based on singular perturbation [16] and a passive approach [17]. Research on null-space interaction control in multi-priority controllers [18], [19] ensures the convergence of task-space errors. The proposed framework by Yu [20] could be extended to the interaction control of slung load transported by multiple aerial vehicles. Ott [21] described Cartesian interaction control and its pros and cons for torque-controlled redundant robots.

Studying the interaction variations in human-like bipedal walking suggests that variable interaction control can improve gait quality and reduce energy loss [22]. Contact between robots and humans allows the human to control the contact forces by adjusting arm stiffness; contact forces can be increased by making the individual's arm stiffer and vice versa [23]. The ability to change the system interaction characteristics based on tasks is one of the key factors for the good performance of biomechanical systems, such as adaptivity and agility. Therefore, variable interaction control has become a popular capability of modern robot interaction operations, which is key to safely completing complex operation tasks.

Robots could automatically change interaction control parameters by interacting with an unknown environment rather than manually adjust the complex parameter values each time. This concept coincides with the idea of RL. The main idea of RL is to find a policy with the highest payoff function to meet current needs through continuous interactive trial and error with the environment. This does not need a priori knowledge, such as a complex model

*Supported by the National Natural Science Foundation of China under Grant 62173037, National Key R. D. Program of China, and State Key Laboratory of Robotics and Systems (HIT). (Corresponding author: Yushu Yu)

Yuting Feng, Chuanbeibei Shi, Jianrui Du and Yushu Yu are with the School of Mechatronical Engineering, Beijing Institute of Technology, China 3220205034@bit.edu.cn; 2535009447@qq.com; dujr0017@163.com; yushu.yu@bit.edu.cn

Fuchun Sun and Yixu Song are with the Department of Computer Science and Technology, Tsinghua University, China fcsun@mail.tsinghua.edu.cn; songyixu@hotmail.com

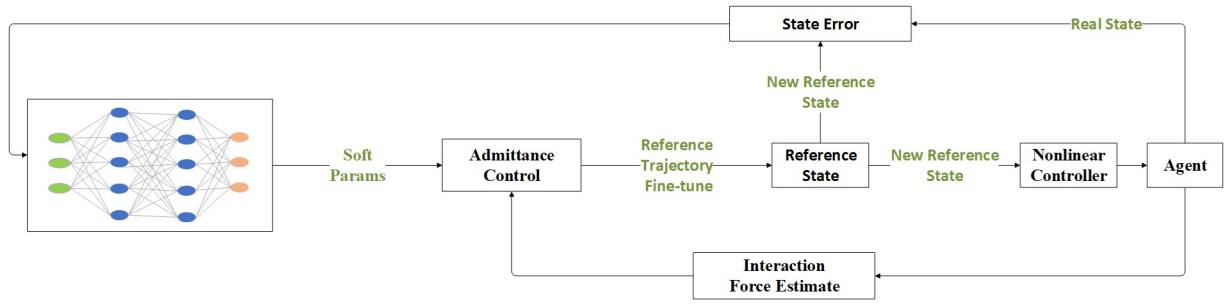

Fig. 2. Overall frame design, where the inputs of the network are the state errors and the outputs are the soft parameters for the admittance control to fine-tune the reference trajectory.

of the controll system, which enables complex autonomous compliant control. Many scholars have studied the use of RL algorithms to learn parameter adjustment policies and dynamically adapt the interaction characteristics of robots.

Dimeas and Aspragathos [3] enhanced the accuracy of robot position control and reduced the energy required through RL. The model-free variable interaction control algorithm and forward neural network as the evaluator has given robots human-like variable interaction control abilities [24]. A model-free RL method, called PI2, adjusts the reference trajectory and interaction parameters simultaneously through path integration for variable interaction control, which is successful in a variety of high-dimension control applications [25]. To get external forces, this paper performs periodic predictions of the dynamic external forces using a force estimation method [26]. Learning from demonstrations not only allows robots to model manipulation tasks but also allows automatic adaptation to unknown situations [27].

Martín [28] studied the effects of different action spaces using deep RL and advocated for variable impedance control in end-effector space as an advantageous action space for constrained contact-rich tasks. An approach proposed for dynamic environments is an adaptation policy that adjusts the control gains of a standard impedance controller to reject disturbances [29]. Many previous interactive control methods could not adapt to multiple scenarios simultaneously or did not consider different UAV platforms.

Our contributions can be summarized as follows:

- We establish a new criterion to evaluate the interaction performance based on energy consumption and state tracking principles.
- We propose a novel policy by combining force estimates and a neural network that can output the variable stiffness and damping parameters, which is trained via RL.
- The proposed model has a certain robustness and can be used across multiple scenes and UAV platforms. Good results are achieved in different interaction force scenarios and dissimilar UAV platforms in the simulation environment.

The remainder of this paper is organized as follows. The background knowledge is described in Section II. The network design and learning framework are outlined in Section III. The results of the simulation environment are shown in Section IV. Finally, the conclusion and future work are given in Section V.

II. BACKGROUND

A. UAV Dynamics Model

The considered UAV is an X-configuration quadrotor, as shown in Fig. 1. A body-fixed reference frame $\boldsymbol{\Sigma}_b$ placed at the UAV's center of mass and an inertia reference frame $\boldsymbol{\Sigma}_i$ are defined. The UAV is an underactuated mechanical system with six degrees of freedom but four independent control inputs. The dynamic equations related to the UAV are inferred by the Newton-Euler formulation as

$$m\ddot{\boldsymbol{p}}_b^b = -m\boldsymbol{S}(\boldsymbol{w}_b^b)\dot{\boldsymbol{p}}_b^b + m\boldsymbol{R}^T\boldsymbol{g} + \boldsymbol{f}_b^b + \boldsymbol{f}_u^b(\cdot) \tag{1}$$

$$\dot{\boldsymbol{R}} = \boldsymbol{R}\boldsymbol{S}(\boldsymbol{w}_b) \tag{2}$$

$$\boldsymbol{I}_b\dot{\boldsymbol{w}}_b^b = -\boldsymbol{S}(\boldsymbol{w}_b^b)\boldsymbol{I}_b\boldsymbol{w}_b^b + \boldsymbol{\tau}_b^b + \boldsymbol{\tau}_u^b(\cdot) \tag{3}$$

where m is the mass of the UAV, $\boldsymbol{I}_b$ represents the inertia tensor of the UAV, $\boldsymbol{p}_b^b \in \mathbb{R}^3$ is the position of the UAV, $\boldsymbol{R} \in SO(3)$ is the rotation matrix representing the attitude of the UAV, $\boldsymbol{w}_b$ denotes the angular velocity of the UAV expressed in $\boldsymbol{\Sigma}_i$, $\boldsymbol{S}(\cdot)$ denotes the skew-symmetric matrix, $\boldsymbol{g} = [0\ 0\ g]^T$ is the gravity vector with $g = 9.81m/s^2$, $\boldsymbol{f}_b^b \in \mathbb{R}^3$ and $\boldsymbol{\tau}_b^b \in \mathbb{R}^3$ are the force and torque input vectors respectively, expressed in $\boldsymbol{\Sigma}_b$, and $\boldsymbol{f}_u^b \in \mathbb{R}^3$ and $\boldsymbol{\tau}_u^b \in \mathbb{R}^3$ denote unknown forces and moments based on the vehicle-aerodynamic and buoyancy effects, imbalances caused by batteries and/or onboard sensors, motion of a robotic arm (or other moving sensors, e.g. a laser scanner on a pan-tilt mechanism) mounted on the aerial platform, parametric uncertainties, wind gusts, flapping dynamics [30], interactions with the environment, etc.

Fig. 3. An example of a UAV putting out a fire

B. Force Estimate

Due to the dynamic formulation of the UAV, we use an estimator of the unmodeled dynamics and external wrench (forces and torques) that acts on the UAV to compensate for some disturbance effects [26]. The expression of the estimate external wrench and unmodeled dynamic $\boldsymbol{r}(t) = [\hat{\boldsymbol{f}_u}^T \ \hat{\boldsymbol{\tau}_u}^T] \in \mathbb{R}^{6\times1}$ in the time domain is defined as:

$$\begin{aligned}\boldsymbol{r}(t) = \boldsymbol{K}_1(\int_0^t -\boldsymbol{r}(\sigma) + \boldsymbol{K}_2(\boldsymbol{q}(\sigma) - \int_0^t (\boldsymbol{r}(\sigma) \\ + \begin{bmatrix} -u\boldsymbol{R}\boldsymbol{i}_3 + m\boldsymbol{g} \\ \boldsymbol{\tau}_b^b - \boldsymbol{S}(\boldsymbol{w}_b^b)\boldsymbol{I}_b\boldsymbol{w}_b^b \end{bmatrix})d\sigma)d\sigma)\end{aligned} \tag{4}$$

where $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are positive definite diagonal matrices, $\boldsymbol{q}(\sigma)$ is the momentum vector of the dynamics system, $\boldsymbol{i}_3 = [0\ 0\ 1]^T$, and u represents the thrust perpendicular to the propeller rotation plane. We use this force estimator to replace the 6D force/torque sensor.

C. Impedance and Admittance Control

In interactive control, the robot can express the impedance ideally as:

$$\boldsymbol{M}\ddot{\tilde{\boldsymbol{p}}}_b + \boldsymbol{D}\dot{\tilde{\boldsymbol{p}}}_b + \boldsymbol{K}\tilde{\boldsymbol{p}}_b = \boldsymbol{r}(t) \tag{5}$$

where $\boldsymbol{M} \in \mathbb{R}^{6\times6}$ is a desired positive definite diagonal inertia matrix, $\boldsymbol{D} \in \mathbb{R}^{6\times6}$ is a desired positive definite diagonal damping matrix, $\boldsymbol{K} \in \mathbb{R}^{6\times6}$ is a desired positive definite diagonal stiffness matrix, and $\tilde{\boldsymbol{p}}_b \in \mathbb{R}^6$ is the errors of position and angle errors.

Compared with the impedance controller, the admittance controller can decouple impedance control from motion control actions to better offset uncertainty in the original control model. The admittance controller could be readily implemented on an existing control loop. During motion, the parameters of the admittance controller can be updated to balance the tracking effect and improve safety.

D. Reinforcement Learning

Control in a continuous action space is difficult for RL, but there have been significant advances using neural networks with RL. The deep deterministic policy gradient (DDPG) algorithm improves on the continuous action domain of the deep q-network (DQN) [31], as it employs an actor-critic architecture and uses two neural networks for each actor and critic to learn a model-free policy. The trust region policy optimization (TRPO) algorithm guarantees monotonic improvements, although it is a policy gradient method. This work selects the proximal policy optimization (PPO) method [32] due to its use for UAV low-level control [33], and because it has outstanding performance in contact environments [28]. While the PPO has similarities to the TRPO and is also a policy gradient method, it is easier to implement and tune.

III. APPROACH

The training for the UAV to appropriately regulate the desired damping and stiffness parameters is based on RL. The RL approach is formalized as a markov decision process(MDP), which is a discrete-time stochastic control. At each time step t, the net observes the current state $\boldsymbol{s}_t$ of the dynamic system and performs an action $\boldsymbol{u}_t$ by adjusting the desired damping and stiffness parameters of the admittance controller, which selects from a set of consecutive actions. After applying these actions, the system arrives at a new state $\boldsymbol{s}_{t+1}$ and obtains a reward r_{t+1}. The policy $\boldsymbol{\pi}$ determines the actions at each state. The $\boldsymbol{\pi}$ is rated by a value function based on the receiving rewards.

Noise processing of UAV sensor data is performed to increase the model robustness. This can alleviate some negative effects caused by unstable sensor data in actual processes. We use a normal distribution function to simulate the sensor noise data which is obtained from real UAVs. To train the network policy, we add a certain range of random noise to the quality and inertia of the UAV. Our model does not need to train for the test scene in the experiment. Instead, a circular trajectory is used during training, and the force is a distribution, such as a cosine function. This is mainly because the distribution of the cosine function is wider and can cover more data than a single type of force, which is true for the scene 1 and scene 2 mentioned later. In addition, noise processing is also performed for the output actions of the network, which could increase its exploration ability.

The PPO algorithm from RL is adopted here. As shown in Fig. 2, the inputs of the network are the error values of 18-dimensional, $\boldsymbol{s}_t = [\tilde{\boldsymbol{p}}\ \dot{\tilde{\boldsymbol{p}}}\ \tilde{\boldsymbol{R}}\ \tilde{\boldsymbol{w}}]$. The inputs include the errors of the position $\tilde{\boldsymbol{p}} \in \mathbb{R}^3$, linear velocity $\dot{\tilde{\boldsymbol{p}}} \in \mathbb{R}^3$, attitude $\tilde{\boldsymbol{R}} \in \mathbb{R}^9$(represented as a rotation matrix), and angular velocity $\tilde{\boldsymbol{w}} \in \mathbb{R}^3$. These errors are between the real values and the reference values. The outputs are a map of the damping and stiffness parameter values of the admittance controller, which are 12-dimensional. The policy network is a fully connected two-layer system, and the outputs of the net are μ and σ. These are then the inputs of Gaussian model to indirectly obtain the compliance control parameters. During training, we add some action noise to increase the exploration ability of the model. Hyperparameters of the network settings are shown in Table I. This paper trains the network over approximately 2000 episodes. Low-level control of the UAV is based on a nonlinear controller [34]. In simulation the network runs at 100 Hz and the dynamics integration is executed at 200 Hz.

TABLE I
HYPERPARAMETERS OF THE PPO ALGORITHM

Batch_size	Hidden_size	Policy_class	Clip_range
512	[64,64]	GaussianMLPPolicy	0.05

IV. SIMULATION AND RESULTS

We perform simulations on the quadrotor dynamical systems and evaluate different scenarios to investigate the model robustness. All network training is finished using the PPO algorithm as the model-free policy optimization method. The scenes include three situations: putting out a fire, fixed thrust, and sliding along a wall. We compare our method with many

other policies that include the min function, max function, mid function, random function, and static random function.

The min function is in a fixed range, and we adopt the minimum parameters at all times regardless of what the environment is. In contrast, the max function is the maximum of the parameters with a fixed range. The mid and rand functions operate similarity with the median and random parameter selection. The static random is a random parameter only in the first step but remains constant thereafter. However, the parameters of the random function are variable in every step. These approaches increase the credibility of the simulation. We call our model policy "net" during the simulations.

We fixed the parameters $\boldsymbol{D}$ and $\boldsymbol{K}$ with a certain range to improve the system stability. As admittance control is similar to a second-order spring-damped system, maintaining stability requires a certain representation of the parameters as:

$$\hat{\boldsymbol{D}} = 2\boldsymbol{\xi}\boldsymbol{\omega} \tag{6}$$

$$\hat{\boldsymbol{K}} = \boldsymbol{\omega}^2 \tag{7}$$

where $\boldsymbol{\omega} \in \mathbb{R}^6$ and $\boldsymbol{\xi} \in \mathbb{R}^6$ are the desired natural frequency and damping of the designed estimator, $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{K}}$ are the main diagonal elements of $\boldsymbol{D}$ and $\boldsymbol{K}$, respectively. The outputs of the "net" are $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$. The limitations of $\boldsymbol{D}$ and $\boldsymbol{K}$ could transfer to the limitations of $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$. The range of the parameters is given in Table II, and the range of ξ_i by [35], which represents the i_{th} element of $\boldsymbol{\xi}$. At the same time, ω_i represents the i_{th} element of $\boldsymbol{\omega}$. We set ω_i by testing the dynamic model stability.

TABLE II
RANGE OF THE NATURAL FREQUENCY AND DAMPING

	ω_i	ξ_i
Range	$[3, 20]$	$(0, 1]$

To ensure the robot can adapt to the environment under interactive operation, we restrict contact forces and the threshold of control gains to ensure safety. Small control gains allow the system to have many desired characteristics, like reduced wear, while high control gains make the system stable. Usually a greater control gain gives additional energy loss, which is also in line with the human variable admittance regulation rules: as compliant as possible, only when the task needs to increase rigidity. That is, we should ensure the tracking accuracy and reduce the values of control gains.

An energy consumption term is used in the cost function to ensure the robot applies variable admittance characteristics. The admittance gain required to complete a task is reduced by punishing the control action. The instantaneous cost function of the admittance control gain is defined as:

$$R_{DK} = -k_{damp}\left\|\hat{\boldsymbol{D}}\right\|^2 - k_{stiffness}\left\|\hat{\boldsymbol{K}}\right\|^2 \tag{8}$$

In addition, to ensure the trajectory tracking while being compliant, it is necessary to limit trajectory errors. The state error reward is defined as:

$$R_{state} = -k_{pos}\left\|\tilde{\boldsymbol{p}}\right\|^2 - k_{vel}\left\|\dot{\tilde{\boldsymbol{p}}}\right\|^2 \tag{9}$$

To maintain stability during flight, the number of collisions and UAV stability are constrained as:

$$R_{stable} = -k_{crash}I_{crashed} - k_{acc}\left\|\ddot{\tilde{\boldsymbol{p}}}\right\|^2 \tag{10}$$

where $I_{crashed}$ is the crashed representation of the UAV. If the UAV crashes, the $I_{crashed}$ is 1; otherwise, it is 0.

Therefore, the final reward is set as:

$$R_t = R_{DK} + R_{state} + R_{stable} \tag{11}$$

A. Scene 1: Put out a Fire

Scene 1 is to put out a fire, which is similar to the environment shown in Fig. 3. We used an impulse function to simulate the scene with a force shown in Fig. 4 Scene 1. The results are given in Table III, which illustrates that the min and net policies are better. However, the reward of the net policy is the best.

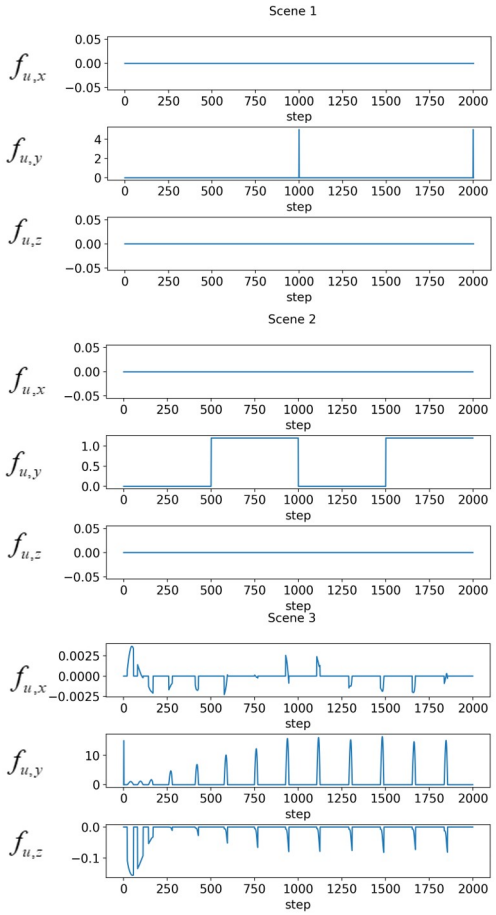

Fig. 4. The contact forces of different scenes. The x-axis represents the time step and the y-axis is contact forces $\boldsymbol{f}_u = [f_{u,x}\ f_{u,y}\ f_{u,z}]$. From top to bottom, the data are from Scene 1, Scene 2, and Scene 3.

TABLE III
POLICY REWARDS

Scene	*Policy*	*Reward*
Scene 1: Putting out a fire	Max	-2956.950
	Mid	-1085.325
	Rand	-1066.797
	StaticRand	-1940.828
	Min	-137.471
	Net	**-119.768**
Scene 2: Fixed thrust	Max	-2972.560
	Mid	-1100.261
	Rand	-1074.317
	StaticRand	-292.382
	Min	-165.726
	Net	**-159.631**
Scene 3: sliding along a wall	Max	-2970.676
	Mid	-1099.199
	Rand	-1132.636
	StaticRand	-2048.502
	Min	-159.280
	Net	**-155.654**

B. Scene 2: Fixed Thrust

Scene 2 is a fixed thrust along the y-axis, where a static forcing function is used in the simulation with details given in Fig. 4 Scene 2. We assume that the force changes at a fixed frequency. We use a fixed force in the y-axis of 1.4 N because a larger external force would decrease the probability that the UAV system could stabilize. The selected force is within the stable threshold but it is not the maximum allowable value. The force image is given in Fig. 4 Scene 2, where the external force is zero in the first 500 steps and becomes 1.4 N from steps 500-1000. Between 1000 and 1500 steps, the force again becomes zero. The force along all other axes remains zero throughout the entire simulation. As shown in Table III, the proposed net policy has the best reward.

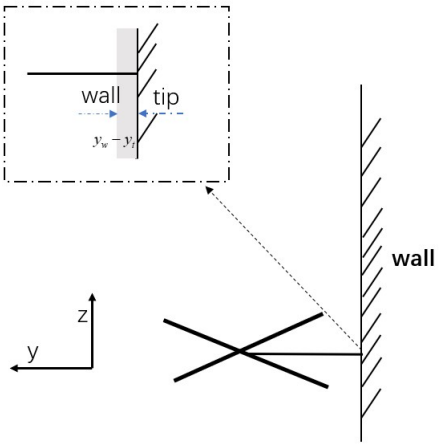

Fig. 5. The end effector of a UAV contacts a surface and slides along one direction. In the dashed box, the dashed arrow on the left points to the surface of the wall, and the dashed arrow on the right points to the tip of a rigid link driven by the UAV. The $y_w - y_t$ represents the deformation distance.

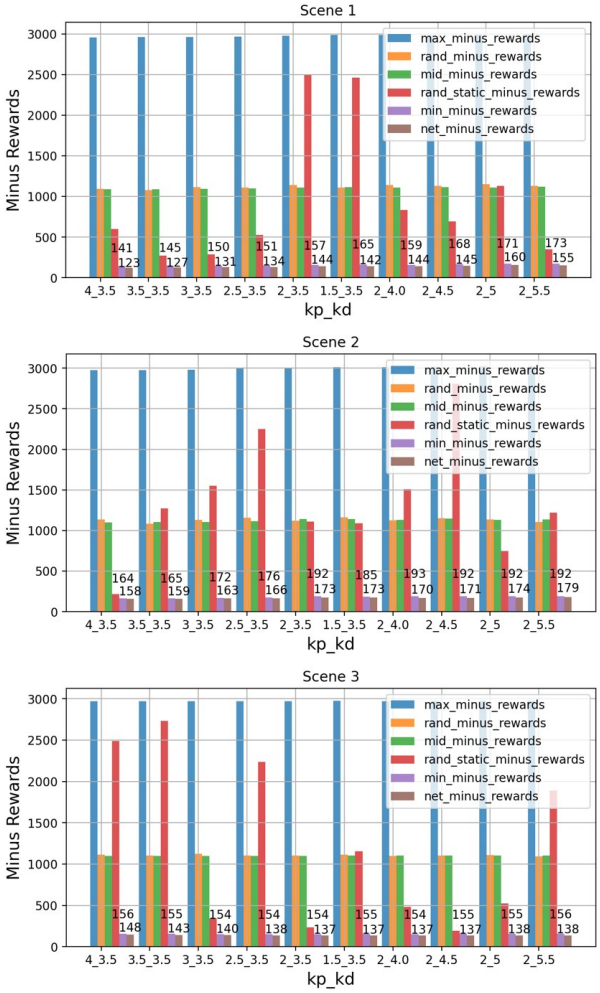

Fig. 6. Illustration of the various low-level controller parameters in the three considered scenarios for different policy tests. From top to bottom, the data are from Scene 1, Scene 2, and Scene 3. The specific values of the min and net policies are marked on the top of the bar chart. The height of the corresponding bar graph is the opposite number of the reward.

C. Scene 3: Sliding Along a Wall

Scene 3 is a UAV that slides along a wall. The external force acting on the tool tip is considered as follows. The interaction forces are from the surface along the positive direction, and the friction forces oppose the direction of the sliding motion on the surface. Here, the positions of the end effector and the wall are given in Fig. 5. The interaction forces from the surface are calculated as:

$$f_{u,y} = k_{wall}(y_w - y_t) \tag{12}$$

where k_{wall} is the elasticity coefficient of the wall, y_t is the position of the end effector along the y-axis, and y_w is the surface position of the wall along the y-axis. In this simulation experiment, $y_w = 0$. The surface friction forces, have a function form of:

$$\boldsymbol{f}_{friction} = -\mu \boldsymbol{v_t} \tag{13}$$

where μ is the coefficient of friction, and $\boldsymbol{v}_t$ represents the velocity of the end effector. In the contact model, the force

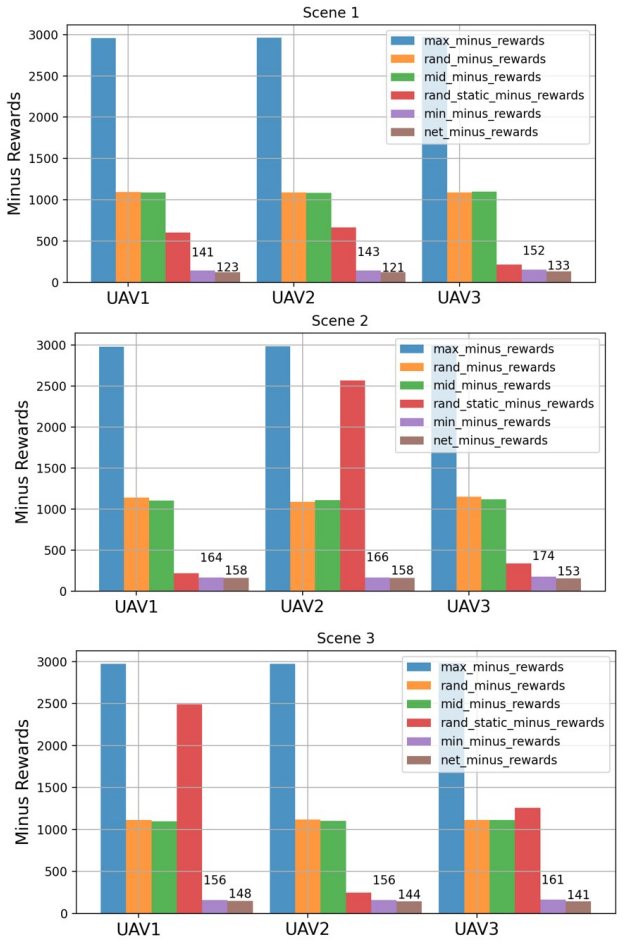

Fig. 7. Comparison of the admittance strategies for different UAV platforms and scenes.

exits only when the end effector tip (hereafter called tip) penetrates to the surface. So the final external forces in this scene are defined as:

$$\boldsymbol{f} = \begin{cases} \boldsymbol{f}_{friction} + [0 \ f_{u,y} \ 0], & f_{u,y} < 0 \\ [0 \ 0 \ 0], & f_{u,y} > 0 \end{cases} \tag{14}$$

We first set the position of the tip to 0 while the target position of the UAV along the y-axis is -0.05 m because the tip should be in contact with the wall. During the initial phase, the tip contacts the wall and quickly bounces off. However, due to the 0.05 m y-axis goal, the tip attempts to contact the wall again and slides upward along it. During the entire simulation, the tip continuously bumps into the wall and bounces off while moving upwards until leaving the wall. The wall is assumed to have an infinite height. Under different strategies, the external forces generated are not completely consistent due to changes in the admittance parameters. The force image of the min policy is shown in Fig. 4 Scene 3. From Table III, the proposed net policy has the best reward performance.

We adjusted the low-level control parameters in the simulation experiments with ten groups of parameters using a nonlinear controller to prove the robustness of the net policy. The results shown in Fig. 6 indicate that the proposed net policy gives the best results. The x-axis represents the parameters of the low-level control k_p and k_d, and the y-axis is the opposite number of rewards. Further, the results indicate that the stability of the net policy is not affected by the low-level control parameters.

We also test using different UAV platforms, as shown in Fig. 7. The x-axis represents the different UAV platforms, and the y-axis is the opposite number of rewards. We change the mass, inertial, and motor parameters of the UAV, as shown in Table IV. The results indicate that the net strategy is still optimal.

TABLE IV
UAV PLATFORM PARAMS

	Mass(kg)	***Inertia***	***Motor (thrust2weight)***
UAV1	1.557	[0.0091,0.0091,0.0141]	1.939
UAV2	0.654	[0.0039,0.0034,0.0063]	2.800
UAV3	0.018	[1.2e-5,1.4e-5,2.93e-5]	1.900

V. CONCLUSIONS AND FUTURE WORK

This paper presents a method for variable interaction control based on RL. We propose an evaluation criterion that considers the reward for integrating energy consumption and trajectory tracking. The parameters of the network were trained based on the proposed reward of admittance control. Compared with a variety of other parameter generation policies, the proposed net policy was optimal. Three possible scenes were designed for the current network, and improved result were obtained by the net policy in the simulations. The adaptation of the net policy was proven by the different parameters of the low-level control. At the same time, we proved that the net policy was optimal for various UAV models. Future work will consider relevant experimental verifications on the visual environment platform and the entity machine.

REFERENCES

[1] J. Luo, E. Solowjow, C. Wen, J. A. Ojea, A. M. Agogino, A. Tamar, and P. Abbeel, "Reinforcement learning on variable impedance controller for high-precision robotic assembly," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 3080–3087, 2019.

[2] M. Ryll, G. Muscio, F. Pierri, E. Cataldi, G. Antonelli, F. Caccavale, D. Bicego, and A. Franchi, "6d interaction control with aerial robots: The flying end-effector paradigm," *The International Journal of Robotics Research*, vol. 38, no. 9, pp. 1045–1062, 2019.

[3] F. Dimeas and N. Aspragathos, "Reinforcement learning of variable admittance control for human-robot co-manipulation," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1011–1016, IEEE, 2015.

[4] M. Bernard, K. Kondak, I. Maza, and A. Ollero, "Autonomous transportation and deployment with aerial robots for search and rescue missions," *Journal of Field Robotics*, vol. 28, no. 6, pp. 914–931, 2011.

[5] T. Bartelds, A. Capra, S. Hamaza, S. Stramigioli, and M. Fumagalli, "Compliant aerial manipulators: Toward a new generation of aerial robotic workers," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 477–483, 2016.

[6] Y. Yu, C. Shi, D. Shan, V. Lippiello, and Y. Yang, "A hierarchical control scheme for multiple aerial vehicle transportation systems with uncertainties and state/input constraints," *Applied Mathematical Modelling*, vol. 109, pp. 651–678, 2022.

[7] Y. Yu, P. Li, and P. Gong, "Finite-time geometric control for underactuated aerial manipulators with unknown disturbances," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 13, pp. 5040–5061, 2020.

[8] V. Lippiello and F. Ruggiero, "Exploiting redundancy in cartesian impedance control of uavs equipped with a robotic arm," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3768–3773, IEEE, 2012.

[9] B. Siciliano and L. Villani, "An inverse kinematics algorithm for interaction control of a flexible arm with a compliant surface," *Control Engineering Practice*, vol. 9, no. 2, pp. 191–198, 2001.

[10] A. Albu-Schaffer and G. Hirzinger, "Cartesian impedance control techniques for torque controlled light-weight robots," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, vol. 1, pp. 657–663, IEEE, 2002.

[11] F. Caccavale, P. Chiacchio, A. Marino, and L. Villani, "Six-dof impedance control of dual-arm cooperative manipulators," *IEEE/ASME Transactions On Mechatronics*, vol. 13, no. 5, pp. 576–586, 2008.

[12] T. Eiband, M. Saveriano, and D. Lee, "Learning haptic exploration schemes for adaptive task execution," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7048–7054, IEEE, 2019.

[13] D. Tsetserukou, N. Kawakami, and S. Tachi, "Obstacle avoidance control of humanoid robot arm through tactile interaction," in *Humanoids 2008-8th IEEE-RAS International Conference on Humanoid Robots*, pp. 379–384, IEEE, 2008.

[14] Y. Li, S. S. Ge, and C. Yang, "Impedance control for multi-point human-robot interaction," in *2011 8th Asian Control Conference (ASCC)*, pp. 1187–1192, IEEE, 2011.

[15] Y. Li, S. S. Ge, C. Yang, X. Li, and K. P. Tee, "Model-free impedance control for safe human-robot interaction," in *2011 IEEE International Conference on Robotics and Automation*, pp. 6021–6026, IEEE, 2011.

[16] A. Albu-Schaffer, C. Ott, U. Frese, and G. Hirzinger, "Cartesian impedance control of redundant robots: Recent results with the dlr-light-weight-arms," in *2003 IEEE International conference on robotics and automation (Cat. No. 03CH37422)*, vol. 3, pp. 3704–3709, IEEE, 2003.

[17] A. Albu-Schäffer, C. Ott, and G. Hirzinger, "A unified passivity-based control framework for position, torque and impedance control of flexible joint robots," *The international journal of robotics research*, vol. 26, no. 1, pp. 23–39, 2007.

[18] H. Sadeghian, L. Villani, M. Keshmiri, and B. Siciliano, "Multi-priority control in redundant robotic systems," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3752–3757, IEEE, 2011.

[19] E. M. Hoffman, A. Laurenzi, L. Muratore, N. G. Tsagarakis, and D. G. Caldwell, "Multi-priority cartesian impedance control based on quadratic programming optimization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 309–315, IEEE, 2018.

[20] Y. Yu, K. Wang, R. Guo, V. Lippiello, and X. Yi, "A framework to design interaction control of aerial slung load systems: transfer from existing flight control of under-actuated aerial vehicles," *International Journal of Systems Science*, vol. 52, no. 13, pp. 2845–2857, 2021.

[21] C. Ott, *Cartesian impedance control of redundant and flexible-joint robots*. Springer, 2008.

[22] J. van den Kieboom and A. J. Ijspeert, "Exploiting natural dynamics in biped locomotion using variable impedance control," in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 348–353, IEEE, 2013.

[23] K. Lee and M. Buss, "Force tracking impedance control with variable target stiffness," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 6751–6756, 2008.

[24] F. Stulp, J. Buchli, A. Ellmer, M. Mistry, E. A. Theodorou, and S. Schaal, "Model-free reinforcement learning of impedance control in stochastic environments," *IEEE Transactions on Autonomous Mental Development*, vol. 4, no. 4, pp. 330–341, 2012.

[25] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal, "Learning variable impedance control," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 820–833, 2011.

[26] F. Ruggiero, J. Cacace, H. Sadeghian, and V. Lippiello, "Passivity-based control of vtol uavs with a momentum-based estimator of external wrench and unmodeled dynamics," *Robotics and Autonomous Systems*, vol. 72, pp. 139–151, 2015.

[27] F. J. Abu-Dakka, L. Rozo, and D. G. Caldwell, "Force-based variable impedance learning for robotic manipulation," *Robotics and Autonomous Systems*, vol. 109, pp. 156–167, 2018.

[28] R. Martín-Martín, M. A. Lee, R. Gardner, S. Savarese, J. Bohg, and A. Garg, "Variable impedance control in end-effector space: An action space for reinforcement learning in contact-rich tasks," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 1010–1017, IEEE, 2019.

[29] W. Zhang, L. Ott, M. Tognon, and R. Siegwart, "Learning variable impedance control for aerial sliding on uneven heterogeneous surfaces by proprioceptive and tactile sensing," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11275–11282, 2022.

[30] S. Omari, M.-D. Hua, G. Ducard, and T. Hamel, "Nonlinear control of vtol uavs incorporating flapping dynamics," in *2013 IEEE/RSJ international conference on intelligent robots and systems*, pp. 2419–2425, IEEE, 2013.

[31] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

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

[33] A. Molchanov, T. Chen, W. Hönig, J. A. Preiss, N. Ayanian, and G. S. Sukhatme, "Sim-to-(multi)-real: Transfer of low-level robust control policies to multiple quadrotors," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 59–66, IEEE, 2019.

[34] D. Mellinger and V. Kumar, "Minimum snap trajectory generation and control for quadrotors," in *2011 IEEE international conference on robotics and automation*, pp. 2520–2525, IEEE, 2011.

[35] T. Wimbock, C. Ott, A. Albu-Schaffer, A. Kugi, and G. Hirzinger, "Impedance control for variable stiffness mechanisms with nonlinear joint coupling," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3796–3803, IEEE, 2008.