CONFIDENTIAL. Limited circulation. For review only.

Feedforward Enhanced Reinforcement Learning Control for a Small Humanoid Robot KHR-3HV*

Yucheng Xin, Linqi Ye, Xueqian Wang*

Abstract— **The humanoid gait generation task of biped robots has been a lasting challenge. To achieve a stable gait in such a complex kinematic system, with the improvement of computing power and algorithms updates, machine learning methods such as reinforcement learning (RL) are widely applied to this topic and made progressive results. RL provides a convenient end-to-end solution and has become the mainstream machine learning method in the field of robotics. However, there are still problems when using traditional RL methods in aspects of similarity to human natural gait, training time, robustness to perturbances, and generalization ability of other tasks. In this paper, we propose an improved framework based on the feedforward enhanced reinforcement learning (FERL) algorithm to efficiently generate a humanoid gait for a small humanoid robot KRH-3HV. In FERL, the control action of biped robots consists of two parts: the reference part and the weighted RL part. In this paper, we introduced prior knowledge that human walking gait exhibits sinusoidal characteristics and designed reference actions by inverse kinematic analysis based on this principle. The weighted RL part is independently set to motivate the biped robot to complete the target task. Providing a reference control signal sequence helps to generate an ideal gait and the action space of the weighted RL is decreased compared with traditional RL so that the training time is reduced while still maintaining certain robustness. By setting up 4 groups of control experiments in a simulation environment, we efficiently generated an effective humanoid gait to complete the target tasks of walking on flat ground and climbing stairs, and proved that the proposed method works well in gait biomimetic similarity, training efficiency, robustness, and generalization ability to other tasks by simple reward function designing.**

I. Introduction

The gait generation of biped robots is one of the most significant topics in the research fields of robotics and biomimetics. With a wide range of applications including military detection, high-risk works, logistics, and more, biped robots are also the foundation for achieving embodied intelligence, and humanoid robots, which provide a stable platform for movement, adapt to complex terrains, and accomplish locomotion tasks.

Biped robots have been developed by many laboratories and institutions around the world for years. The ASIMO robot launched by HONDA in 2000 is absolutely a milestone, with 57 degrees of freedom (DOF), it can run at the max speed of 9 kilometers per hour and accomplish complex tasks like going up and down stairs[1]. In 2015, the Valkyrie, which was designed by NASA, was published. The robot was initially planned for the missions of space exploration. It was able to independently complete the tasks such as opening and closing doors with its 44 degrees of freedom[2]. In recent years, the Cassie biped robot which is launched by Agility Robotics in 2017 has become one of the most popular research biped robots with its unique biomimetic structure and excellent function[3]. In 2023, the Atlas, which was developed by Boston Dynamics, has already been able to backflip, roll forward, pick non-standardized objects, and climb steps with load. Meanwhile, Tesla showcased its own humanoid robot named Optimus at Tesla AI Day in 2023, which could grab small objects, and transport cargoes with a depth camera. All of these investments in biped robot have made great success and is predicted to generate great value in the future.

The control algorithms for biped robots also have made a lot of progress in recent years. Traditional control algorithms such as Zero-Moment Point (ZMP)[4], Whole-Body Control (WBC)[5], and Hybrid Zero Dynamics (HZD)[6] have worked well. In traditional control theory, the models of biped robots are usually abstracted as a linear inverted pendulum (LIP)[7]. Some algorithms were inspired to achieve certain natural kinematic mechanisms of humanoid gait. Constrained by energy and computing power issues, although traditional control algorithms work well in some cases, it is still hard for them to excavate the full potential of biped robots. With the improvement of computing power and the emergence of artificial neural networks, a large number of machine learning and deep learning algorithms have been widely applied in the field of biped robot control[8][9]. The most mainstream method currently can be reinforcement learning.

Reinforcement learning is good at solving model-free optimization problems. By setting appropriate reward functions to achieve an action policy mapping from observation states to action decisions. In 2007, the team of Jungho Lee in KAST used reinforcement learning for walking pattern generation and got great simulation results with abstract the robot as a planar inverted pendulum model[10]. In 2017, the team of Nicolas Heess in DeepMind used reinforcement learning to train robots to cross various complex terrains including hurdles, platforms, gaps, slalom walls, and more[11]. In 2021, the team of Jonah Siekmann at OSU launched a framework that made a biped robot climb stairs without vision information by generating random stair terrains in the training period[12]. In the early works of this team in 2020, the robot had already generated a stable gait with a reduced order model mechanism which could be well

*This work was supported by the National Natural Science Foundation of China No. 92248304 and No. 62003188.

Yucheng Xin is with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China (e-mail: usfinea@163.com).

Linqi Ye is with the Institute of Artificial Intelligence, Shanghai University, Shanghai, China (e-mail: yelinqi@shu.edu.cn).

Xueqian Wang is with the Center of Intelligent Control and Telescience of the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China (corresponding author provide phone: 0755-26031120; e-mail: wang.xq@sz.tsinghua.edu.cn).

Manuscript 416 submitted to 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO). Received August 15, 2023.

combined with high-dimension information like visual data to complete some specific tasks[13].

In addition to traditional reinforcement learning methods, imitation learning can be regarded as an extension of reinforcement learning. After generating reference gaits or control signals as ideal data, the robot is trained to imitate the reference pattern by constructing a reinforcement learning network to maximize the reward which is set to account for the differences between the policy outputs with the reference actions. In 2021, the team of Zhongyu Li in UCB generated a reference gait by hybrid-zero dynamics methods and worked well in sim-to-real in Cassie to complete tasks like fast walking, turning around, and recovering from foot sliding[14]. In 2022, the team of Heyuan Yao in PKU used a world model mechanism to augment the reference data and enhanced the robustness and generalization in simulation[15].

Reinforcement learning provides an effective end-to-end method that can be used in almost all kinds of biped robots, and it appears high robustness because of its exploring mechanisms especially when handling random perturbation or non-standardized situations. But it also requires a long training time to explore all possible states in real locomotion and insufficiency to generate prescriptive ideal locomotion patterns. So that traditional reinforcement learning methods can well resolve tasks like walking or running, but it's hard to generate a natural humanoid gait or generate a gait that is easy to employ in real robots. And it's difficult to design reward functions when dealing with complex tasks. On the other hand, imitation learning can generate a natural humanoid gait as expected by training to converge to reference gait but with low robustness in special situations and high dependence on the reference dataset so that it could only be applied on robots that the reference data already exist or easy to generate.

To deal with the problems above, the feedforward enhanced reinforcement learning (FERL) is proposed[16]. In FERL, the robot control action is divided into two parts: the reference part and the weighted RL part. By involving a reference action, the FERL combined it with the weighted RL output. In this paper, we propose an improvement framework based on FERL. We introduced prior knowledge that human walking gait exhibits sinusoidal characteristics[17]. And the reference action is designed based on this principle of the features of real human gait. It provided an ideal locomotion pattern, although it may not control the robot to complete the target task well, the weighted RL is used to fine-tune the control signals of the robot based on the reference signals to keep it stable and finish the task. The decreased action space, due to the involvement of the weighted RL, results in reduced training time and still offers a certain level of robustness. Without dependence on the specific reference dataset to imitate, the proposed method can be used in almost all biped robots. For verification, we used the KHR-3HV robot which was developed by KONDO in a simulation environment. And by setting up 4 groups of controlled trials, we motivate the robot to complete the tasks of walking on flat ground stably and climbing stairs, and prove that the proposed method could successfully generate an ideal gait with more similarity to natural humanoid posture, less training time, high-robustness and great generalization ability for other tasks than traditional reinforcement learning.

II. METHODS

A. Framework

The framework proposed is depicted in Figure 1. The first step is to construct an ideal gait generator by inverse and forward kinematic analysis and set it as a reference control action. In this paper, we introduced prior knowledge that human walking gait exhibits sinusoidal characteristics and constructed the reference gait in the framework according to this principle[17]. The reference gait could not be so perfect for the target task. It just confines an ideal locomotion pattern. The second step is to build an appropriate deep reinforcement learning algorithm to train the robot to finish the target task. This step is set to improve the effects of the reference gait. The third step is to combine the reference signals and the reinforcement learning signals to the final control signals with a weighting mechanism that is involved in restricting the action spaces of the RL to decrease the time-consuming of the training procedure. In summary, we have constructed a framework using a weighted RL affected by a reference signal. The observation states in the RL part come from the sensors recording the posture and motion information, and the reward is set to complete the target tasks.

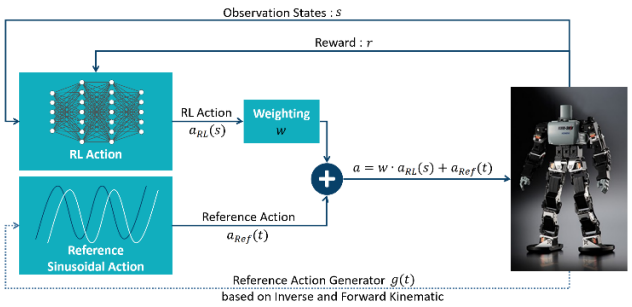


Figure 1. Framework

Although the reference gait may not be so fit for the target task, it still contains ideal locomotion pattern information which means every reference signal is close to the optimal solution in action space for the present task in part. And weighted reinforcement learning is set to fine-tune the reference signals to make the robot converge to the optimal gait and complete other related tasks including keeping stability and rejecting disturbances.

B. Inverse Kinematics Analysis

The forward and inverse kinematic analysis is used to generate a reference control signal which leads robots to accomplish the target task as much as possible. As for a walking task to generate a humanoid gait, it is necessary to analyze the mechanical structural characteristic of the robot and derive the inverse kinematic function of its body part of lower limbs.

In this paper, we use the KHR-3HV robot, which is shown in Figure 2, as a sample to validate the effectiveness of the framework. For the lower limbs of its body part, it has 5 degrees of freedom (DOF) in each leg. The servos in lower limbs are installed on the hips, thighs, shanks, ankles, and feet of the robot, and respectively result in rotation in directions of roll, pitch, pitch, pitch, and roll.

Manuscript 416 submitted to 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO). Received August 15, 2023.

Figure 2. KHR-3HV Robot

For the convenience of analysis, we define this standing state of the joints' angle position which is also shown on the right side of the Figure 2 as a zero-point position. The positive angle values represent the rotation towards outside for the roll direction, and the negative angle values represent the rotation towards backward for the pitch direction.

The inverse kinematic analysis aims to obtain all of the joints' position trajectories in the lower limbs to make the end terminal move to the target position. Considering the particularity of mechanical structure and ideal walking motion, the end target of inverse kinematics can be set as the junction midpoint between the ankle and shank part in each leg. So that the inverse kinematic problem is abstracted as a two-linking rod model. We can define directions of the coordinate system which marks forward as the positive direction of the X-axis, leftward as the positive direction of the Y-axis, and upward as the positive direction of the Z-axis. Two representative states of the lower limbs for the resolution of inverse kinematics have been shown in Figure 3.

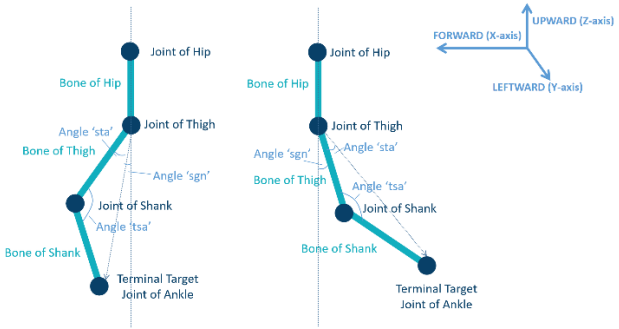


Figure 3. Lower Limbs' Inverse Kinematic Solution

The joint points in Figure 3 also represent the connection part of related articulations. To maintain the foot at a given position and keep level, it could be easy to derive the joint angle values with a given target position within a reachable range as follows:

$$\theta_{hip} = -\arctan\frac{p^{Y}_{hip,target}}{p^{Z}_{hip,target}}, \tag{1}$$

$$\theta_{thigh} = -\hat{\theta}_{sta} + \mathrm{sgn}(p^{X}_{ankle}) \cdot \hat{\theta}_{sgn}, \tag{2}$$

$$\theta_{shank} = 180° - \hat{\theta}_{tsa}, \tag{3}$$

$$\theta_{ankle} = -\theta_{thigh} - \theta_{shank}, \tag{4}$$

$$\theta_{foot} = -\theta_{hip}, \tag{5}$$

where, θ_{hip}, θ_{thight}, θ_{shank}, θ_{ankle}, θ_{foot} represent the real joint angle to control the terminal limbs to reach the target position. p^{X}_{ankle} means projection on the Y-axis of the position of the ankle joints in body space. $p^{Y}_{hip,target}$ represents projection on the Y-axis of the direction vector from the position of the hip to the target position, $p^{Z}_{hip,target}$ is for the same reason representing projection on the Z-axis. Each of $\hat{\theta}_{sgn}$, $\hat{\theta}_{sta}$ and $\hat{\theta}_{tsa}$ indicates the intersection angular between two direction vectors, which is explained in Figure 3, and it's easy to obtain their values through the Cosine Theorem. All of the position vectors are defined in body space rather than in world space. The angular constraint in formula (5) ensures that the foot can always provide a supporting surface parallel to the body.

In this way, the inverse kinematic model of KHR-3HV has been resolved so that it can be used to control leg postures arbitrarily by controlling the terminal target position.

C. *Sinusoidal Reference Gait*

From the description above, it's clear that the posture of the legs can be designed by adjusting the terminal target position in inverse kinematic models and thereby controlling the robot to generate a desired gait as a reference. So that it is particularly necessary to design an ideal terminal target trajectory to generate a reference gait that can walk steadily and conform to characteristics of the real humanoid gait. By analyzing the characteristics of the natural walking gait of humans, we can learn that the human gait contains a lot of sinusoidal features in the world coordinate we defined in the last section[17]. In the forward direction, the feet exhibit alternating linear movement. In the upward direction, the feet change height in a sinusoidal frequency. In the lateral direction, the feet change the projection distance away from the body with a sinusoidal motion shape, and it results in a sinusoidal waving of the body relative to the supporting surface composed of feet. We can draw the motion curves of the robot's lower limbs and body based on three coordinate planes in Figure 4.

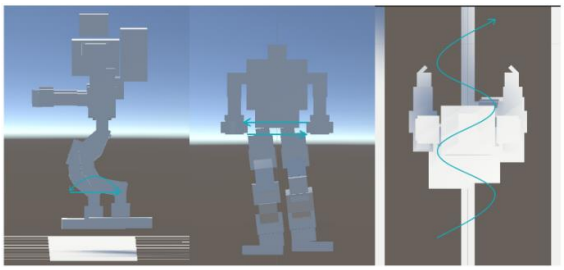


Figure 4. Sinusoidal Gait Generated by Inverse Kinematic

Through the analysis above, we can obtain the motion trajectory of the inverse kinematic terminal target, which is located in the junction of ankle joints, in the body coordinate system for a walking gait. The whole period of walking gait could be divided into four parts. In the first part, the center of the body needs to swing above the supporting leg with both of two feet in the ground. In the second part, the behind foot raises and moves forward with the front foot moving backward at a stable height. Meanwhile the center of the body waves to the center of the body trajectory. The rest two parts are the mirror action of the previous two parts, the center of the

body move the above of supporting leg and finish the exchange of the front and rear position of both feet.

The parameters that determine a walking gait include the step time period, step distance, initial foot position, step height, and the swing angular of the body which can all be set by adjusting the positions of inverse kinematic terminal targets. The reference gait has already been generated in such a way. We can obtain detailed locomotion trajectories of each joint to participate in subsequent training. It should be noted that the reference gait cannot be so perfect for the walking task, but it must contain enough information to represent the ideal gait pattern like the step distance, step height, and time period for movement. So that we can fine-tune the control action to obtain a better gait by using reinforcement learning.

D. Reinforcement Learning

Once a reference humanoid gait for walking is generated, we are equivalent to obtaining a set of joint control signal data which can also be used in imitation learning. The general purpose of reinforcement learning (RL) is to train a mapping from observation states to action signals used in robot control. The imitation learning is also aiming to train such a mapping, but varies from RL, IL is set to imitate an action trajectory to minimize differences from the given reference actions over time or state domain changes.

TABLE I. PARAMETERS OF PPO

Hyperparameters	***Network Setting***
fixed timestep : 0.005s	hidden units : 512
max step in episode : 5000 steps	num layers : 3
batch size : 2048	normalize : true
buffer size : 20480	vis_encode_type : simple
learning rate : 0.0003	memory : null
beta : 0.005	goal_conditioning_type : hyper
epsilon : 0.2	deterministic : false
lambda : 0.95	
num_epoch : 4	

In this paper, our goal is also to obtain a mapping from observation states to robotic control action signals. The core point is that the outputs of the policy network in RL are not the final control signals to the robot. The control signals consist of two parts: the signals from the reference gait we generated from inverse and forward kinematic analysis, and the weighted output from the RL network. In the RL mechanism, the agent will explore randomly in action space to converge to the optimal solution for every observation state. The exploration mechanism helps to consider all kinds of possibilities for the agent to obtain better robustness. However, the more action parameters, the larger the action space, and the longer the training time. For the robot itself, the reference gait signal has already provided a solution that may work well in action space with a mapping from the state last time step to the action decision what is happening. So that the ideal humanoid gait can be fine-tuned based on the reference gait signals, and that's how the weighted RL works. By weighting the output of RL, the action space of random exploration has been reduced and results in decreasing in training time. In this paper, we use the PPO as the RL algorithm part. The detailed parameters of PPO are shown in Table I.

TABLE II. OBSERVATION STATES

Parameters	***Meanings***	***Dimensions***
θ_i	Angular of each joint	10
q_b	Rotation quaternion of the body	4
w_b	Angular velocity of the body	3
v_b	Velocity of the body	3

- Observation states

The observation states are set as showcased in Table II. θ_i represents the angular of joints. The lower limbs which are needed to control contain 10 joints in total, with 5 joints in each leg. (q_b, w_b, v_b) represents the states of the body in locomotion.

- Actions

The outputs of the deep RL network are used to form the control signals for each joint in the lower limbs, and there are 10 joints to control in a whole. Because the weighting mechanism is involved, the final control signal for each joint consists of two parts from the RL and the inverse analysis of the reference gait.

- Reward Design

The reward function can be constructed according to different task requirements. For every task we want to complete in this paper, the reward function can be combined as:

$$r = r_l + r_s + r_a \tag{6}$$

where the r_l represents the living reward for promoting the agent not to trigger the termination conditions like falling, the r_s represents the ability to keep stable when locomotion, the r_a is set to motivate the agent to take action decisions to complete the specific tasks.

In this paper, we are going to validate that the framework works better in generating gait that is more similar to the natural state of humans in morphology, and it has better training efficiency, robustness, and generalization ability to other tasks than traditional RL methods. So that we design two tasks of walking on flat ground and climbing the stairs. The reward functions of each task are designed as showcased in Table III:

TABLE III. REWARD FUNCTION DESIGNS

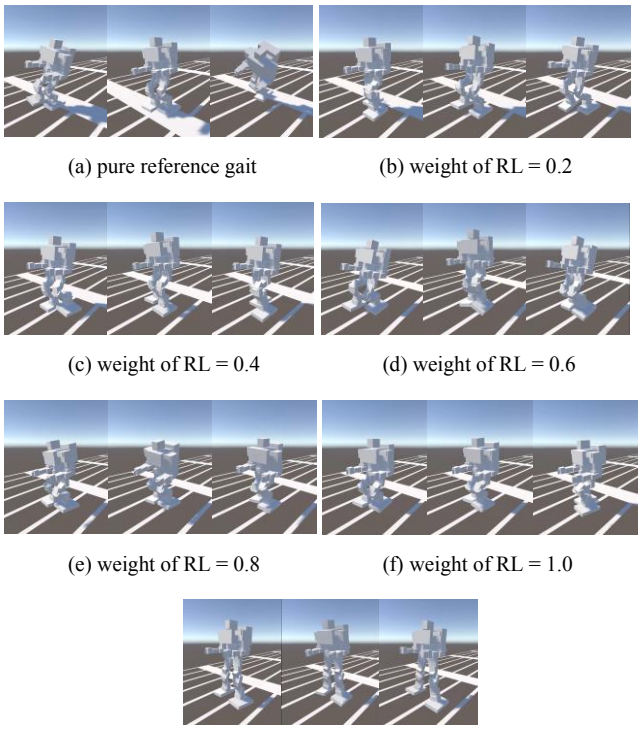
Reward Parameters	***Task I: Walking on flat ground***	***Task II: Climbing the stairs***
r_l	1	-2
r_s	$-0.01 \cdot (\lvert\theta_{roll}\rvert + \lvert\theta_{pitch}\rvert + \lvert\theta_{yaw}\rvert)$	
r_a	$3 \cdot e^{-\lvert v_b^X - 0.3\rvert}$	$e^{-\lvert v_b^X - 0.3\rvert} + p_{body}^Y + p_{body}^X$

where the $(\theta_{roll}, \theta_{pitch}, \theta_{yaw})$ represent the deviation angles of the robot body along three coordinate axes in world space, v_b^X represents the velocity of the body part in forward direction, (p_{body}^X, p_{body}^Y) represents the horizontal movement distance and, the height of the robot to indicate the degree of climbing stairs.

To this end, we can describe the action decisions of the robot we expect in different tasks. For the task of walking on flat ground, we hope the robot to stay long alive so that the r_l is set as a positive value. But for the task of climbing stairs, we hope the robot climbs to the target position as soon as possible so that we set the r_l as a negative value. In order to maintain the stability of the robot in locomotion, the deviation angles of the trunk are hoped to be as small as possible which determines the form of r_s. As for specific motion requirements, the robot is supposed to move forward and raise its height during climbing stairs, so we design to reward the locomotion in the forward direction at a target velocity and upward direction in the task of climbing stairs.

III. Results

In this section, we are to show that the proposed framework performs well in generating a humanoid gait. For this purpose, we designed 4 sets of experiments. In these experiments, we compare the proposed method with the traditional RL method in aspects of the similarity to human walking gait, time consumption efficiency in training, and robustness to perturbance. The traditional RL is defined that the joint control signals of the robot are directly equivalent to the outputs of the RL networks without the feedforward. Meanwhile, by realizing the new task of climbing stairs, we demonstrate that the framework also has great adaptability and migration ability for other tasks.



(a) pure reference gait (b) weight of RL = 0.2

(c) weight of RL = 0.4 (d) weight of RL = 0.6

(e) weight of RL = 0.8 (f) weight of RL = 1.0

(g) pure RL

Figure 5. Gait Generated with Different RL Weighting Value

All experiments in this paper are conducted in a simulation environment. We use KHR-3HV, which is developed by KONDO, as a robot model and ML-Agents package for the simulation and reinforcement learning. The fixed time step in simulation is 0.005s. To focus on the locomotion of the robot's lower limbs, we will fix the posture of the robot's upper limbs in experiments in simulation.

A. Biomimetic Similarity

In this section, we designed a comparative experiment to validate that the proposed framework can generate a gait, which is more similar to the nature gait of humans in the task of walking on flat ground. We compared the effectiveness of the gait generated after training of 4 million steps with different weighting values of weighted RL when the reference signals are involved. And we set a control group with traditional RL and no reference signals involved. The final effect of generating gait is showcased in Figure 5.

The Figure 5(a) shows the pure reference gait generated by inverse kinematic analysis without any RL signals involved. It is clear that the pure feedforward may not complete the walking task well, but it does provide information on gait patterns, and the target gait can be adjusted based on this reference. Figure 5(b) to 5(f) show the gait when the reference signals are applied and the weighting value of RL is set from 0.2 to 1. Figure 5(g) shows the gait generated by the traditional RL method, which means the weighting value of RL is set as 1, and no reference signals are used.

From the effects displayed through gaits above, when reference signals are applied, and the weight of RL is within a small range, the robot can successfully generate a humanoid gait. As the weight of RL increases, the exploration space for action decisions of RL is increasing, and the framework will gradually change to the traditional RL method. So that when the weight of RL is set to big such as shown in Figure 5(e), the framework can still generate a gait to complete the task of walking to the target position, but the posture of the robot is so far away from how the real persons like and the locomotion of robot may contain a lot of shaking or motions that are difficult to implement onto the real machine.

The results indicate that we can enable the robot to generate a better humanoid gait by adjusting the weight value of RL to complete the target task in our framework.

B. Learning Efficiency

The framework we proposed can greatly improve the training speed and efficiency. In the experiment mentioned in the previous section III-A, we compared the effects of generated gait with different weight values of RL.

Analyze the training speed and results under these several conditions as shown in the Figure 6, the traditional RL method can be represented by the black line. Similar to the last section, the robot performs well in terms of convergence speed and training effectiveness when reference signals are involved, and the weight of RL is within a small range. When the weight value of RL is set as 0.2, 0.4, or 0.6, the convergence speed to 95% of the maximum value and the maximum reward value are both higher than the values of the traditional RL method. When the weight value of RL is set as 0.2, it takes 0.9 million steps to converge to 95% of the maximum value, but it needs

1.8 million steps for the traditional RL method to achieve the same effect. However, as the weight value of RL increases, the performance becomes worse until it approaches the traditional RL method. This result also indicates that the proposed method can improve training efficiency and effectiveness while generating better humanoid gait by adjusting the weight value of the RL part.

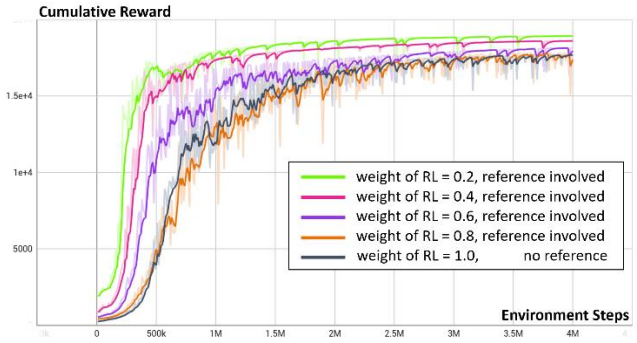
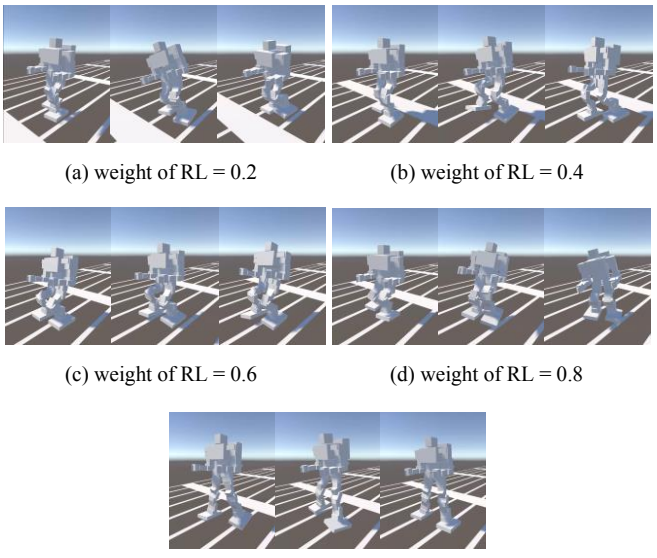


Figure 6. Cumulative Reward Curve while Training

C. *Walking Robustness*

Among various algorithms for biped robots, an advantage of machine learning algorithms lies in their great ability to resist perturbance. The robustness is also an important feature when designing algorithms. To validate the robustness of the proposed method, we introduced random perturbation to the robot within the simulation environment. In each cycle of alternating motion of the robot's feet, the trunk of the robot was subjected to random perturbations along three directions of axes of the world coordinates, ranging from 10N to 300N at a random moment in each cycle. The target task is still same as the section III-A to walk on a flat ground. The gait generated and the variations of the training reward curve are illustrated in Figure 7 and Figure 8.



(a) weight of RL = 0.2 (b) weight of RL = 0.4

(c) weight of RL = 0.6 (d) weight of RL = 0.8

(e) traditional RL

Figure 7. Gait Generated with Random Perturbation

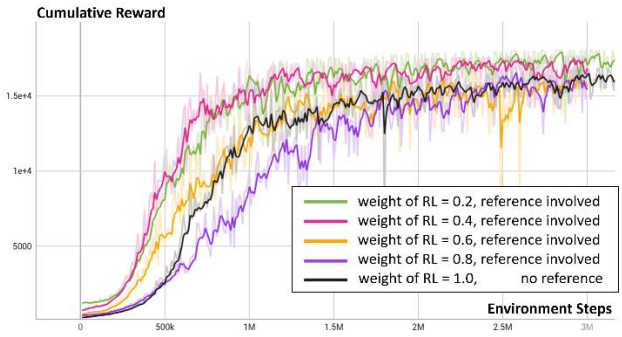


Figure 8. Cumulative Reward Curve with Perturbation

In Figure 7, it can be observed that the phenomenon of experiments is similar to results in section III-A that we can generate a humanoid gait when the weight value of RL is in a certain low range. In almost all cases, robots can still recover to a stable state after being disturbed. However, when the weight value of RL is set too large, problems such as the generated motion, like random shaking, being detached from reality may also occur. And although the action space of joints is reduced because the mechanism of weighted RL is involved, the maximum value of the reward curve does not decrease, as shown in Figure 8. It means that our method with weighted RL can maintain great robustness for robot control.

D. *Generalization to Other Task*

The generalization ability for other tasks is also an important indicator to testify to the effectiveness of an algorithm. The three parts of the experiments above require the robot to complete the tasks of walking on flat ground. In this section, we will drive the robot to complete the task of climbing stairs with the same algorithm.

In this task, the reference gait is the same as the previous task of walking on flat ground, consisting of sinusoidal shape of foot movement and swing of the body part which originated from analysis of human natural locomotion posture. The only difference is the setting of the reward function which is shown in Table III. The training results are shown in Figure 9.

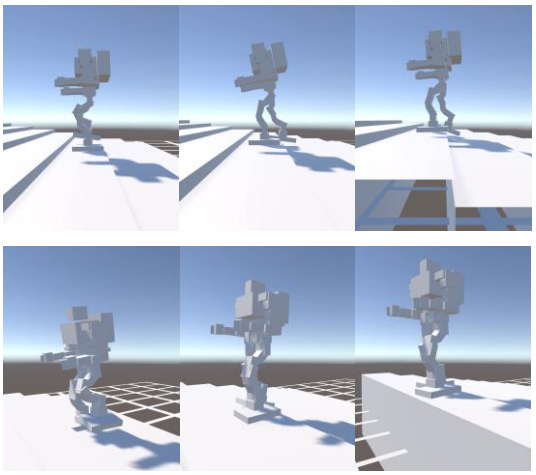


Figure 9. The Robot Climbing Stairs

As shown in Figure 9, it is found that when the weight value of RL is set as 0.2, the robot has already been able to complete climbing stairs with a natural gait after training of 1 million

steps. Compared to this, when we use the traditional RL method, it is still hard to achieve an ability to climb stairs after training of 3 million steps. The result also indicates that the framework we proposed performs well in the generalization of other tasks such as climbing stairs. It helps robots more efficiently generate a great humanoid gait that is capable of completing the target task.

IV. Conclusion

To solve the problems of similarity to human natural gait, training efficiency, robustness to perturbations, and generalization ability of other tasks when using reinforcement learning to generate the humanoid gait of biped robots, we have proposed an improvement framework based on the existing algorithm of FERL. In this framework, the final control action of each joint in the robot consists of two parts: the reference action and the weighted RL action. We introduced prior knowledge that human walking gait exhibits sinusoidal characteristics and designed reference action by inverse kinematic analysis based on this principle. The reference action provides effective information about gait patterns to make gait more humanoid. The weighted RL provides a robust end-to-end method to achieve tasks better by adjusting locomotion posture. With the reference action promoting the generation of a gait closer to real humans, we designed 4 sets of experiments to validate the effectiveness of the framework by requiring the robot to complete tasks of walking on flat ground and climbing stairs. By experimental comparison and verification, the robots can generate a humanoid gait that adapts to the target task to walk on flat ground more efficiently and retain great robustness with a simple reward function design. Finally, the framework is used to enable the robot to accomplish the task of climbing stairs, and it proves the great generalization ability for other tasks of this framework. In the future, we will try to apply the proposed method to the real robot of KHR-3HV.

References

[1] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki and K. Fujimura, “The intelligent ASIMO: system overview and integration,” IEEE/RSJ International Conference on Intelligent Robots and Systems, Lausanne, Switzerland, 2002, pp. 2478-2483 vol.3, doi: 10.1109/IRDS.2002.1041641.

[2] N. A. Radford, P. Strawser, K. Hambuchen, J. S. Mehling, W. K. Verdeyen, A. S. Donnan, J. Holley, et al. “Valkyrie: Nasa's first bipedal humanoid robot,” Journal of Field Robotics 32, no. 3 (2015): 397-419.

[3] Z. Xie, P. Clary, J. Dao, P. Morais, J. W. Hurst and M. Panne, “Iterative Reinforcement Learning Based Design of Dynamic Locomotion Skills for Cassie,” ArXiv abs/1903.09537 (2019): n. pag.

[4] S. Kajita, et al. “Biped walking pattern generation by using preview control of zero-moment point,” 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), Taipei, Taiwan, 2003, pp. 1620-1626 vol.2, doi: 10.1109/ROBOT.2003.1241826.

[5] O. Khatib, L. Sentis, J. Park, et al. “Whole-body dynamic behavior and control of human-like robots,” International Journal of Humanoid Robotics, 2004, 1(01): 29-43.

[6] Y. Gong and J. Grizzle, “One-Step Ahead Prediction of Angular Momentum about the Contact Point for Control of Bipedal Locomotion: Validation in a LIP-inspired Controller,” 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 2832-2838, doi: 10.1109/ICRA48506.2021.9560821.

[7] L. Ye, X. Wang, H. Liu and B. Liang, “The Simplest Balance Controller for Dynamic Walking,” 2022 IEEE International Conference on Robotics and Biomimetics (ROBIO), Jinghong, China, 2022, pp. 1793-1800, doi: 10.1109/ROBIO55434.2022.10011831.

[8] Y. Chun, J. Choi, I. Min, M. Ahn and J. Han, “DDPG Reinforcement Learning Experiment for Improving the Stability of Bipedal Walking of Humanoid Robots,” 2023 IEEE/SICE International Symposium on System Integration (SII), Atlanta, GA, USA, 2023, pp. 1-7, doi: 10.1109/SII55687.2023.10039306.

[9] Z. Xie, G. Berseth, P. Clary, J. Hurst and M. Panne, “Feedback Control For Cassie With Deep Reinforcement Learning,” 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018, pp. 1241-1246, doi: 10.1109/IROS.2018.8593722.

[10] J. Lee and J. H. Oh, “Biped walking pattern generation using reinforcement learning,” 2007 7th IEEE-RAS International Conference on Humanoid Robots, Pittsburgh, PA, USA, 2007, pp. 416-421, doi: 10.1109/ICHR.2007.4813903.

[11] N. Heess, Dhruva, S. Sriram, J. Lemmon, et al. “Emergence of Locomotion Behaviours in Rich Environments,” arXiv, abs/1707.02286.

[12] J. Siekmann, K. Green, J. Warila, A. Fern and J. Hurst, “Blind Bipedal Stair Traversal via Sim-to-Real Reinforcement Learning,” arXiv, abs/2105.08328 (2021): n. pag.

[13] K. Green, Y. Godse, J. Dao, R. L. Hatton, A. Fern and J. Hurst, “Learning Spring Mass Locomotion: Guiding Policies With a Reduced-Order Model,” in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 3926-3932, April 2021, doi: 10.1109/LRA.2021.3066833.

[14] Z. Li, et al., “Reinforcement Learning for Robust Parameterized Locomotion Control of Bipedal Robots,” 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 2811-2817, doi: 10.1109/ICRA48506.2021.9560769.

[15] H. Yao, Z. Song, B. Chen and L. Liu, “ControlVAE: Model-Based Learning of Generative Controllers for Physics-Based Characters,” ACM Transactions on Graphics. 41. 1-16. 10.1145/3550454.3555434.

[16] L. Ye, X. Wang and B. Liang, “Realizing Human-like Walking and Running with Feedforward Enhanced Reinforcement Learning,” in International Conference on Intelligent Robotics and Applications (ICIRA), 2023.

[17] S. Wu, C. Wang, L. Ye, X. Wang, H. Liu and B. Liang, “Quasi-static Walking for Biped Robots with a Sinusoidal Gait,” 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), Mexico City, Mexico, 2022, pp. 849-856, doi: 10.1109/CASE49997.2022.9926469.