actuators

MDPI

Article

A Novel Approach for Train Tracking in Virtual Coupling Based on Soft Actor-Critic

Bin Chen [1,2,3,*], Lei Zhang [1,2,3], Gaoyun Cheng [2,3], Yiqing Liu [2,3] and Junjie Chen [4]

[1] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China; 20111082@bjtu.edu.cn
[2] Traffic Control Technology Co., Ltd., Beijing 100070, China; gaoyun.cheng@bj-tct.com (G.C.); yiqing.liu@bj-tct.com (Y.L.)
[3] National Engineering Research Center of Rail Transportation Operational and Control System, Beijing 100044, China
[4] School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China; junjiec@tsinghua.edu.cn
* Correspondence: chenbin@bjtu.edu.cn

Abstract: The development of virtual coupling technology provides solutions to the challenges faced by urban rail transit systems. Train tracking control is a crucial component in the operation of virtual coupling, which plays a pivotal role in ensuring the safe and efficient movement of trains within the train and along the rail network. In order to ensure the high efficiency and safety of train tracking control in virtual coupling, this paper proposes an optimization algorithm based on Soft Actor-Critic for train tracking control in virtual coupling. Firstly, we construct the train tracking model under the reinforcement learning architecture using the operation states of the train, Proportional Integral Derivative (PID) controller output, and train tracking spacing and speed difference as elements of reinforcement learning. The train tracking control reward function is designed. Then, the Soft Actor-Critic (SAC) algorithm is used to train the virtual coupling train tracking reinforcement learning model. Finally, we took the Deep Deterministic Policy Gradient as the comparison algorithm to verify the superiority of the algorithm proposed in this paper.

Keywords: train tracking; virtual coupling; reinforcement learning; Soft Actor-Critic

Citation: Chen, B.; Zhang, L.; Cheng, G.; Liu, Y.; Chen, J. A Novel Approach for Train Tracking in Virtual Coupling Based on Soft Actor-Critic. *Actuators* **2023**, *12*, 447. https://doi.org/10.3390/act12120447

Academic Editor: Keigo Watanabe

Received: 30 October 2023
Revised: 22 November 2023
Accepted: 28 November 2023
Published: 1 December 2023

1. Introduction

Urban transit systems are a cornerstone of modern urban planning and development. They make cities more livable, equitable, and economically vibrant places by improving the environment and overall urban sustainability. As the size of cities increases, subway lines become more complex, and train operation control tasks are arduous. This will lead to a waste of resources. Virtual coupling in urban subway systems provides solutions to a wide range of challenges, including optimizing train formation, reducing energy consumption, enhancing passenger experience, adapting to variable demand, and ensuring safety. It offers a more responsive and adaptable approach to subway operations, ultimately benefiting both passengers and subway operators. Train tracking control is a crucial component in the operation of virtual coupling trains. It plays a pivotal role in ensuring the safe and efficient movement of trains within the train and along the rail network, and ensures that trains are accurately positioned, aligned, and coordinated during coupling and throughout their journey.

Train tracking control is a crucial component in the operation of virtual coupling trains. It plays a pivotal role in ensuring the safe and efficient movement of trains within the train and along the rail network and ensures that trains are accurately positioned, aligned, and coordinated during coupling and throughout their journey. In virtual coupling, multiple trains form a train formation in which back trains track front trains and there should be a method used to control back trains to front ones. At present, many traditional

methods, such as Model Predictive Control (MPC) and Proportional Integral Derivative (PID) control, are used in the train tracking problems in virtual coupling. However, the control performance of these methods relies on the manual setting of feedback parameters and has the disadvantage of too frequent acceleration and deceleration switching, which cannot guarantee the comfort of passengers. If we can find a method so that the parameters of the control method can be adjusted according to the control effect, the existing train tracking control method can have a better control performance.

In this study, we proposed a novel approach for train tracking in virtual coupling based on Soft Actor-Critic (SAC), which belongs to the reinforcement learning (RL) area. The SAC tends to be sample-efficient, meaning it can learn high-quality policies with relatively fewer samples. And, it is based on the maximum entropy reinforcement learning framework. It employs a maximum entropy policy to balance exploration and exploitation, enhancing the algorithm's robustness and performance. The SAC is used to optimize the parameters of PID by taking the output of PID and the train motion status as the states and the parameters of PID as actions. Then, PID controls the operation of the back train. It is the first time that the SAC has been used in the Virtual Coupling of urban railway transportation. We set up two trains in the Virtual Coupling in this study.

This study is organized as follows: Section 2 presents a literature review about train tracking control problems for railway train operation, especially in Virtual Coupling. Section 3 introduces the train tracking control algorithm in virtual coupling based on the SAC. Section 4 carries out the simulation of the train tracking methods based on SAC proposed in this paper and discusses the results based on real data. Section 5 summarizes the work in this paper and introduces the research work to be performed in the future.

2. Literature Review

In 1999, virtual coupling was first proposed by Bock et al. [1] as well as the new train operation mode, and in 2000, Bock et al. [2] introduced the methods to design and develop the virtual coupling system, which uses train-to-train communication to replace the physical coupler links between train carriages, so that adjacent trains can operate cooperatively and shorten the train tracking interval. Other technical details such as system framework, communication, positioning, etc., are explained in detail in references [3], [4], and [5], respectively.

With the continuous development of train-to-train communication, train positioning, and other technologies, more and more scholars carry out research on the virtual coupling operation mode and train tracking control. Cao et al. [6] proposed a train operation control method based on local leader–follower to establish the virtual coupling implementation scheme, and then a controller used for train tracking was proposed on the basis of parameter identification of the nonlinear virtual coupled train operation process [7]. Lin et al. [8] studied the velocity and constrained tracking control problem to propose a distributed cooperative tracking control algorithm for virtual coupling trains. For high-speed railway virtual coupling, Liu et al. [9] designed a method based on Pontryagin's maximum principle to solve the optimal control problem considering constraints of safe spacing, operation limits, and train dynamic performance. Luo et al. [10] presented a distributed adaptive model predictive control system to coordinate the driving of each unit train in Virtual coupling. Wang et al. [11] illustrated the virtual coupling by a cooperative game model and solved it based on the particle swarm optimization algorithm. Chen et al. [12] proposed an iterative learning control scheme aiming to track the desired reference displacement and velocity in which more general nonlinear uncertainties are imposed on the dynamic model of the train, and that would be closer to reality. Felez et al. [13] proposed a robust MPC method by defining a robust controller into the MPC to control train tracking. A similar work is proposed by Su [14].

The reinforcement learning methods provide a new way for optimized control in many other areas. Wu et al. [15] first investigated the problems of optimal false data injection attack scheduling and countermeasure design for car-like robots in the framework of deep

reinforcement learning. In the area of data-based secure control, the Soft Actor-Critic is used to solve the secure control problem [16]. As for the urban railway, He et al. [17] combined the convolutional neural network with long short-term memory to construct the CNN-LSTM hybrid model from the perspective of trajectory prediction, which can obtain the features both in space and time from measurement data at the same time. They also combined LSTM with the Kalman filter to obtain the long-term dependencies [18]. Huang et al. [19] proposed a cooperative tracking control based on a consensus algorithm and artificial potential field theory to realize the train tracking within a distance range. For the displacement-speed trajectory tracking of the automatic train control system with unknown parametric/nonparametric uncertainties and speed constraints, Li et al. [20] proposed a constrained spatial adaptive iterative learning controller. Zhou et al. [21] proposed an improved disturbance observer-based control method to ensure that the automatic train control (ATC) system can operate safely and control accurately even when the train is affected by uncertainties. Wang et al. [22] introduced deep learning into train tracking to calculate the reference speed trajectory according to the real-time driving condition.

3. Methodology

In this section, we introduce the basic principle of the SAC method used in this article. Then, the train tracking dynamic model is built, which is used as the environment for the reinforcement learning architecture. At last, the train tracking control algorithm based on SAC is illustrated.

3.1. Soft Actor-Critic

SAC [23] is a RL algorithm developed based on the idea of maximum entropy. It uses a randomly distributed policy function (Stochastic Policy) and is an off-policy, Actor-Critic algorithm. It is most similar to other RL algorithms. The difference is that while SAC optimizes the strategy to obtain higher cumulative returns, it also maximizes the entropy of the strategy. SAC has excellent performance in various commonly used benchmarks and real robot control tasks, and its performance is stable and has strong anti-interference ability.

The purpose of standard reinforcement learning is to find the policy that maximizes the expected sum of rewards:

$$\pi^* = \arg\max_{\pi} \sum_t \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t)] \tag{1}$$

where π is the policy in the SAC, s_t is the state at time t, a_t is the action at time t, $r(s_t,a_t)$ is the reward gained by action a_t under state s_t, and ρ_{pi} is the distribution of agent's state s_t and action a_t under policy π.

After introducing the maximum entropy, the target policy of the reinforcement learning can be written as Equation (2):

$$\pi^* = \arg\max_{\pi} \sum_t \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t) + \alpha H(\pi(\cdot|s_t))] \tag{2}$$

where α is a hyperparameter called the temperature parameter, which determines the relative importance of the entropy term against the reward, and thus controls the stochasticity of the optimal policy; $H(\pi(\cdot|s_t))$ is the entropy of $\pi(\cdot|s_t)$ calculated by Equation (3).

$$H(\pi(\cdot|s_t)) = E[-\log \pi(\cdot|s_t)] \tag{3}$$

The action value function of SAC is defined by Equation (4).

$$Q(s_t,a_t) = E_{s_{t+1}\sim\mathcal{D}}[r(s_t,a_t) + \gamma V^\pi(s_{t+1})] \tag{4}$$

where $\mathcal{D}$ is the distribution of previously sampled states and actions or a replay buffer.

The state value function is defined by Equation (5).

$$V(s_t) = E_{a_t \sim \pi}[Q(s_t, a_t) - \alpha \log \pi(\cdot|s_t)] = E_{a_t \sim \pi}[Q(s_t, a_t) + H(\pi(\cdot|s_t))] \tag{5}$$

The loss function of the two q-critic networks can be calculated as Equation (6).

$$\begin{aligned} &J_Q(\theta) = \mathbb{E}_{(s_t, a_t, s_{t+1}) \sim D, a_{t+1} \sim \pi_\phi} \\ &\left[\frac{1}{2}(Q_\theta(s_t, a_t) - (r(s_t, a_t) + \gamma(Q_\theta(s_{t+1}, a_{t+1}) - \alpha \log(\pi_\phi(a_{t+1}|s_{t+1})))))^2\right] \end{aligned} \tag{6}$$

The loss function of the actor-network can be calculated as Equation (7).

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}, \varepsilon \sim \mathcal{N}}\left[\alpha \log \pi_\phi\left(f_\phi(\varepsilon_t; s_t) \mid s_t\right) - Q_\theta\left(s_t, f_\phi(\varepsilon_t; s_t)\right)\right] \tag{7}$$

where $\mathcal{N}$ is the standard normal distribution and f is the probability density function of policy π.

And lastly, the loss function of the temperature parameter α is shown as Equation (8).

$$J(\alpha) = \mathbb{E}_{a_t \sim \pi_t}[-\alpha \log \pi_t(a_t \mid \pi_t) - \alpha \mathcal{H}_0] \tag{8}$$

The process of the SAC algorithm is shown as Algorithm A1 in Appendix A.

3.2. Train Tracking Model Based on Reinforcement Learning

For train tracking in the virtual coupling, it is necessary to ensure that the distance between the two trains is not too large, otherwise the transportation efficiency of the virtual coupling will become low. It is also necessary to ensure that there is a sufficient safe distance between the two trains to ensure the safe operation of the trains in the virtual coupling.

Figure 1 shows the scenario of train tracking in the virtual coupling. In the virtual coupling, the back train (Train 2 shown in Figure 1) changes its position $p_{t,2}$ and speed $v_{t,2}$ at time t by controlling the output of its traction system based on the position $p_{t,1}$ and speed $v_{t,1}$ of the front train (Train 1 shown in Figure 1), which makes the distance Δp_t between Train 2 and Train 1 within a reasonable range. The positions and speeds of Train 1 and Train 2 have the following relationship.

$$\Delta p = p_1 - p_2 - L \tag{9}$$

$$\Delta v = v_2 - v_1 \tag{10}$$

where L is the length of Train 1.

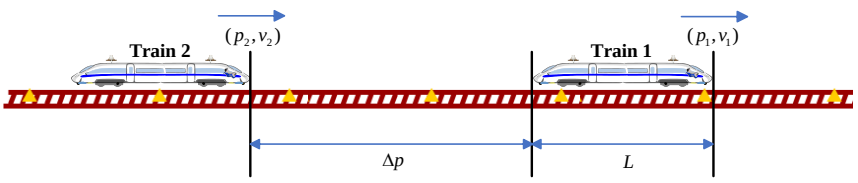

Figure 1. The scenario of train tracking in virtual coupling.

Since the position and speed of Train 1 is the reference target of Train 2, we take Train 2 as the control object and here give its dynamic model as described in our previous paper [24].

PID control is widely used in urban rail transit train operation control due to its simple, real-time, robust, and adjustable characteristics. At the beginning of the operation, there is an initial distance Δp_0 between train 2 and train 1. If the speed of train 2 is controlled to be as close as possible to the speed of train 1, Δp will remain close to Δp_0. Therefore, the

speed of train 2 can be the control target of PID control. Set Δv_t as the difference between v_1 and v_2 at time t ($v_i = v_{t,i}$, for $i \in \{1, 2\}$).

$$\Delta v_t = v_{t,2} - v_{t,1} \tag{11}$$

The PID control variable Δu at time t can be calculated by Equation (12) [25].

$$\Delta u = K_P(\Delta v_t - \Delta v_{t-1}) + K_I \Delta v_t + K_D(\Delta v_t - 2\Delta v_{t-1} + \Delta v_{t-2}) \tag{12}$$

where K_P, K_I, and K_D denote the coefficients for the proportional, integral, and derivative terms, respectively.

Although the PID algorithm is widely used, its parameters need to be finely adjusted. Reinforcement learning learns how to make optimal decisions in different situations through the continuous interaction between the agent and the environment. This article uses the reinforcement learning method to learn the three coefficients of the PID algorithm.

The elements of the RL architecture are described as follows.

Agent: we take the PID controller as the agent of the RL structure, because it can change the operation states of the train by changing the traction or breaking forces of the train.

Action space: Since the RL method is used to learn the three coefficients of the PID algorithm, the action a_t of the RL should be a three-tuple representing the three coefficients.

$$a = \{a_P, a_I, a_D\} \tag{13}$$

It is better to limit the value of action to $[-1, 1]$ as shown in Equation (14), and we transfer a_P, a_I, a_D to K_P, K_I, K_D by Equation (15).

$$a_j \in [-1, 1], \ for\ j\ \in \{P, I, D\} \tag{14}$$

$$K_j = K_{j,min} + \frac{(a_j + 1)(K_{j,max} - K_{j,min})}{2}, \ for\ j\ \in \{P, I, D\} \tag{15}$$

where $K_{j,max}$ is the maximum of K_j and $K_{j,min}$ is the minimum of K_j for $j \in \{P, I, D\}$.

State space: We take the output u of the PID controller as one of the state elements. The speed v_t, acceleration acc_t, and position p_t of the back train at time t are also elements of the state, since they represent the operation characteristics of the train. The differences of the speed Δv_t and the difference of the position Δp_t between the trains in virtual coupling at time t are also in the state of the RL structure. With the limitations on actual train operation conditions shown in Equation (17), the state at time t can be described as Equation (16).

$$s_t = (u_t, acc_t, v_t, p_t, \Delta v_t, \Delta p_t) \tag{16}$$

$$\begin{cases} u_t \in [0, 1] \\ acc_t \in [acc_{min}, acc_{max}] \\ v_t \in [0, v_{max}(t)] \\ p_t \in [0, P_L] \\ \Delta v_t \in [\Delta v_{min}, \Delta v_{max}] \\ \Delta p \in (0, P_L] \end{cases} \tag{17}$$

where the range of the acceleration of the train is $[acc_{min}, acc_{max}]$, and P_L is the length of the whole railway.

Environment: The agent outputs an action and then the environment returns a new state as well as a reward. In this work, the environment should construct the relevance of the output of the PID controller, the state of train's operation and the distance between the two trains in virtual coupling. So, we take the train dynamics model and the train tracking model as the environment of the RL architecture.

Reward: The purpose of train tracking is to keep the distance and the difference of speed between the two trains in virtual coupling in a small range, which means that the distance and the difference of speed should be smaller. Then, we can define the reward $r(t)$ at time t as Equation (18). As Δv_t and Δp_t decrease, $r(t)$ will increase. Therefore, reinforcement learning will learn in the direction where both Δv_t and Δp_t decrease.

$$r(t) = \frac{1}{1+\eta_1|\Delta v_t|} + \frac{1}{1+\eta_2|\Delta p_t|} \tag{18}$$

where η_1 and η_2 are the adjusting coefficients that prevent the reward value from being too large.

Note that Δp_t should always be a positive number since the trains can not collide. The penalty item of the reward is shown as Equation (19).

$$r(t) = -C \ \ if \ \Delta p_t < 0 \tag{19}$$

where C is a large positive number.

3.3. The Train Tracking Control Method Based on SAC

Figure 2 illustrates the architecture of the train tracking control method based on SAC. The SAC generates action a_t and transfers it to (K_P, K_I, K_D). Then, the PID controller generates u by receiving (K_P, K_I, K_D) and derives the train movement under its dynamics. After every step is executed, the operation data are stored in the Replay Memory for the training of SAC. Also, the Replay Memory will be gradually updated by the operation data.

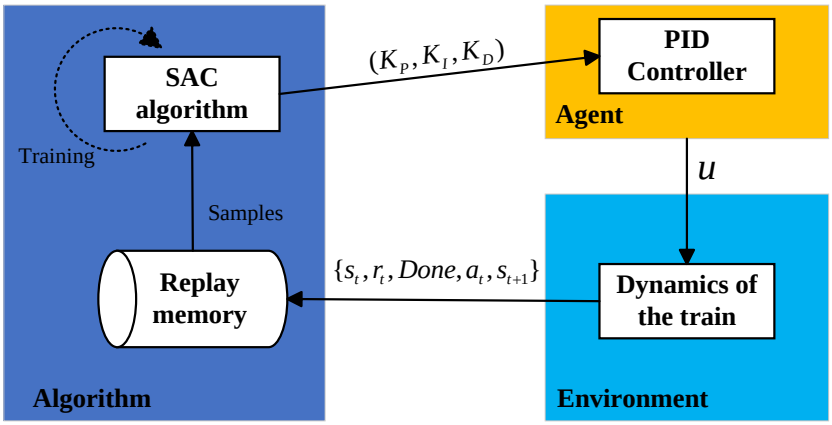

Figure 2. The architecture of the train tracking control method based on SAC.

4. Case Study

To verify the superiority of this work, we took the DDPG algorithm as the comparison algorithm applied to the virtual coupling train tracking control along with the SAC, which is based on the real subway in Beijing. The parameters of the train and the information of the track are similar to our previous paper [24]. Other parameters used in this paper are shown in Table 1.

For the two reinforcement learning methods, we conducted 10,000 episodes of training, respectively. The changing trend of the rewards, the loss, and the train tracking states of virtual coupling are used to illustrate the simulation.

Table 1. Parameters of the PID control and virtual coupling.

Parameter Names	Parameters
$L\ /m$	92
$\Delta p_0/m$	5.92
$\{K_{P,min}, K_{P,max}\}$	$\{0, 5\}$
$\{K_{I,min}, K_{I,max}\}$	$\{0, 1.5\}$
$\{K_{D,min}, K_{D,max}\}$	$\{0, 1.5\}$
η_1	1000
η_2	1000
C	100

Figures 3 and 4 illustrate the changing trend of rewards in two algorithms. Figure 3 shows that the reward of SAC changes a lot before the 4000th episode, and after that, the fluctuation of the original curve decreases and the smoothed curve becomes flat. However, the reward of DDPG shows a downward trend as the training episodes increase, and there is no sign of convergence as shown in Figure 4. The SAC has better performance in this paper than DDPG from the aspect of changing trend of the rewards.

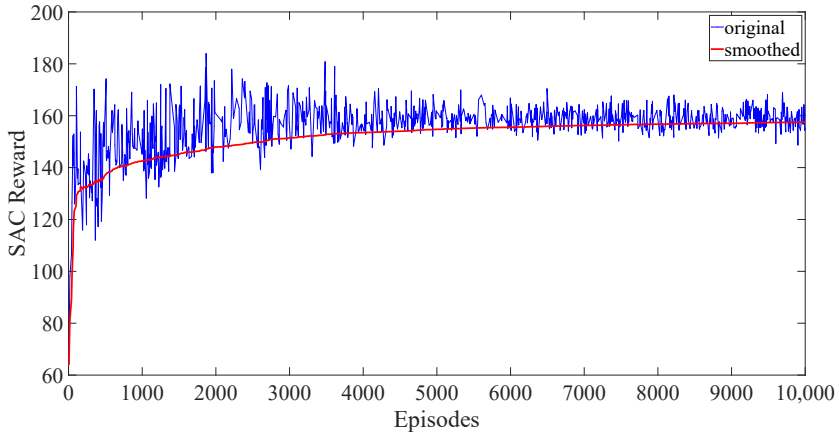

Figure 3. Rewards of SAC.

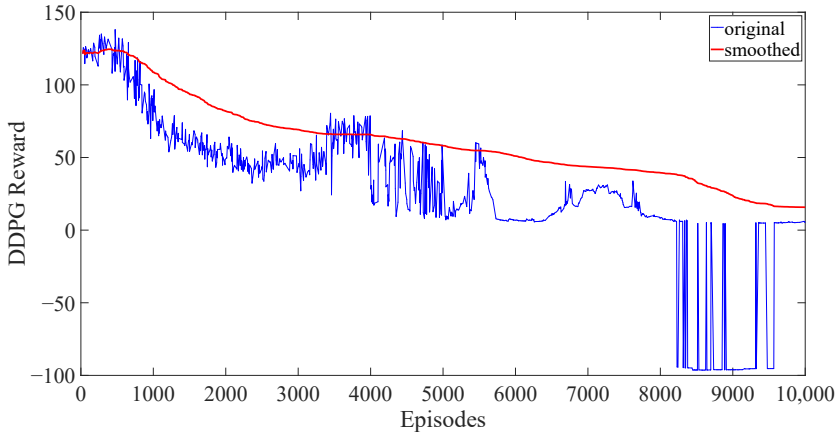

Figure 4. Rewards of DDPG.

Figure 5 shows the changing trend of training losses of SAC and Figure 6 shows that of DDPG. The loss of SAC increases before the 2000th episode, and then decreases until it is stable after the 5000th episode, as shown in Figure 5. It shows that the SAC algorithm keeps learning before the 5000th episode, and then it reaches a stable result for the whole RL progress. As shown in Figure 6, the loss of DDPG does not have a gradual trend. Instead, it

suddenly changes dramatically after 8000 episodes of training. This shows that the learning efficiency of DDPG is very low and its training is unstable.

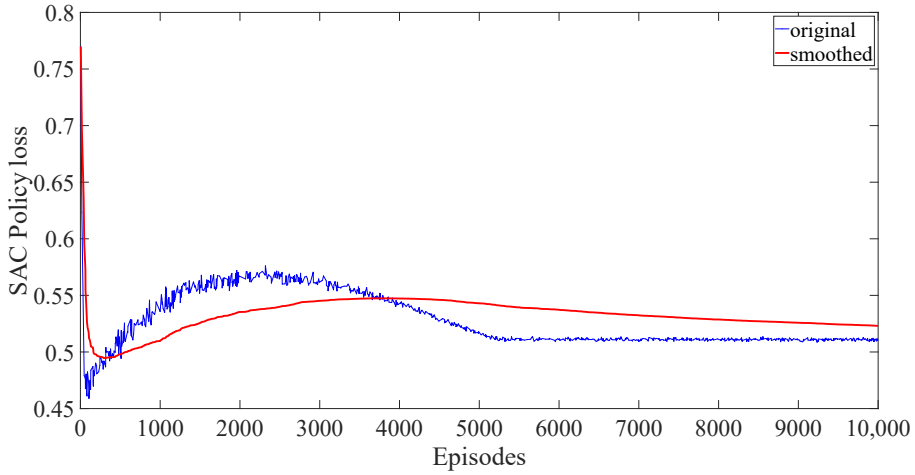

Figure 5. Policy loss of SAC.

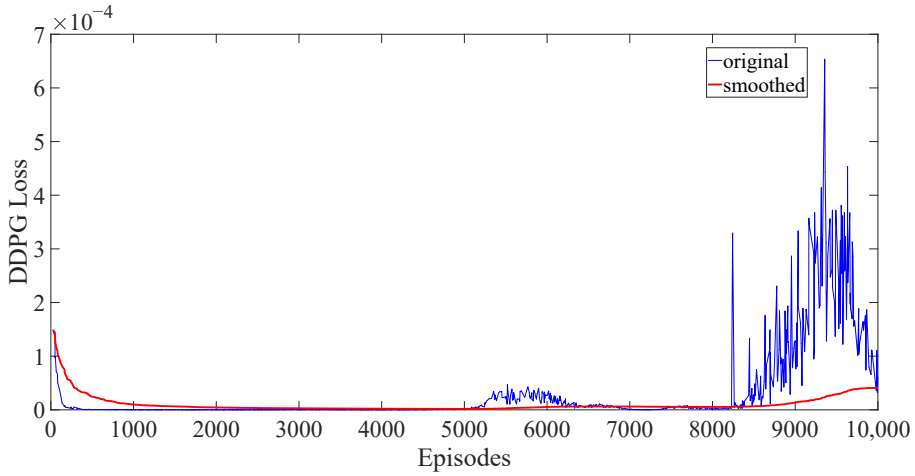

Figure 6. Loss of DDPG.

We can conclude that the DDPG algorithm does not meet the demand of the train tracking control, as its reward does not converge and it has lower learning efficiency than SAC, the SAC performs better than DDPG, and the reward can reach a stable level through the above analysis. For the virtual coupling, the distance and the difference of speed between two trains in virtual coupling are crucial indicators, which are analyzed as follows.

Figure 7 shows the operation profiles of the front train (Train 1) and the back train (Train 2). The trajectories of these two trains are almost similar to each other. The difference in positions and speeds are shown in Figure 8 and Figure 9, respectively. The initial distance between two trains is 5.92 m and from Figure 8 we can see that during the whole operation process, the distance between two trains is between 5.88 m and 5.98 m, and the maximum of Δp is 0.6 m. The two trains in the Virtual Copuling did not collide, which ensures the operation safety in train tracking. The distance between the two trains is not much different from the initial distance. As for the difference of speed, known as Δv, the maximum of $|\Delta v|$ is within 0.15 m/s, as shown in Figure 9, which means that the train tracking efficiency is improved by this algorithm.

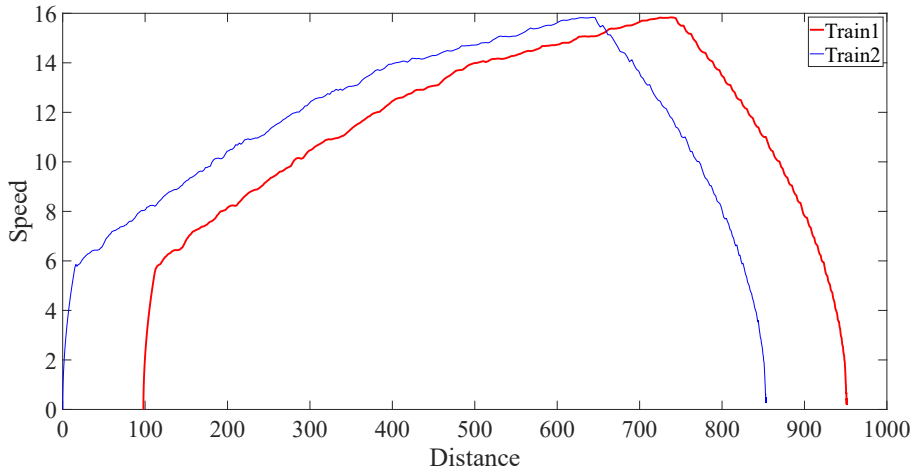

Figure 7. Operation profiles under SAC.

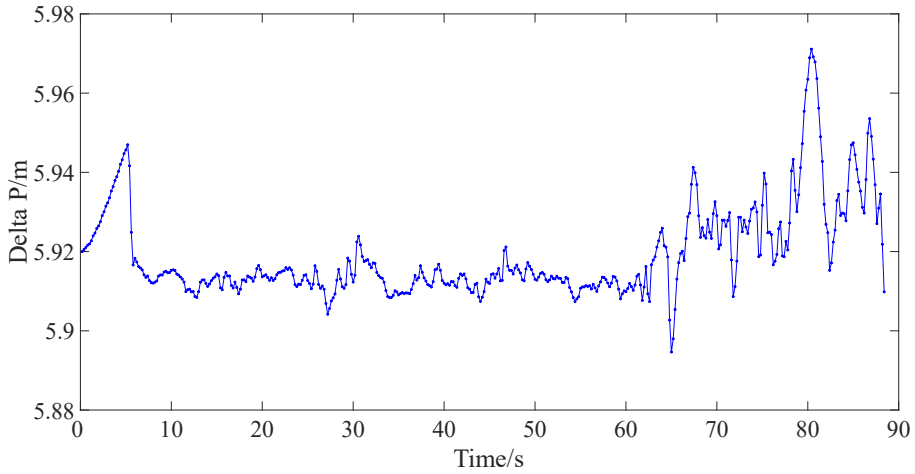

Figure 8. Difference of positions under SAC.

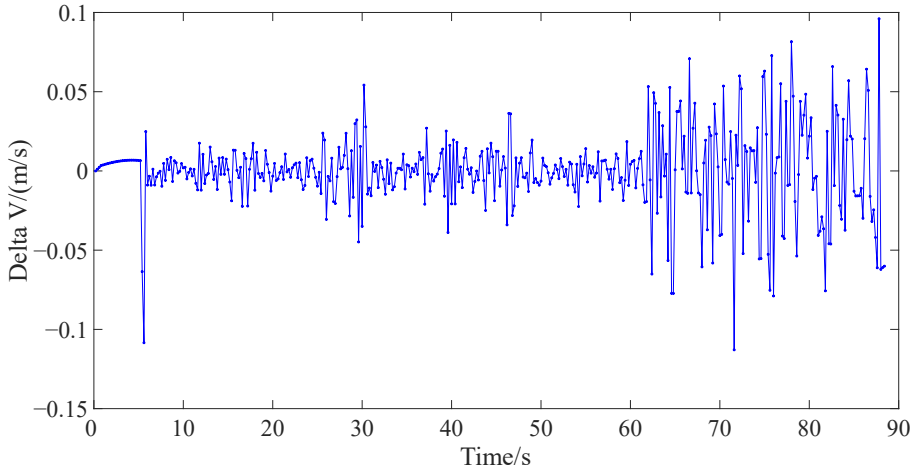

Figure 9. Difference of speeds under SAC.

From the above analysis, it can be seen that under the control method combining SAC and PID, the operation of the two trains in the virtual coupling always maintains a highly consistent state, and the trajectories of these two trains are almost similar to each other. The difference in positions and speeds stays within a small range.

5. Conclusions

The development of virtual coupling technology provides solutions to the challenges faced by urban rail transit systems. Train tracking control is a crucial component in the operation of virtual coupling which plays a pivotal role in ensuring the safe and efficient movement of trains within the train and along the rail network. In order to ensure the high efficiency and safety of train tracking control in virtual coupling, this paper proposes an optimization algorithm based on Soft Actor-Critic for train tracking control in virtual coupling. First, the train tracking model in virtual coupling is constructed including the relationship between the position and speed among trains in virtual coupling, as well as the PID controller. Then, the reinforcement learning model based on SAC for train tracking control is constructed. The train tracking control reward function is designed through the distance and speed difference of the trains. The SAC algorithm is used to train the train tracking reinforcement learning model. The experimental results show that the proposed train tracking control algorithm based on SAC can improve the train tracking efficiency and ensure the safety of virtual coupling.

The proposed method in this article can provide a new way for train tracking control technology in virtual coupling. Combining with traditional control methods can help adjust parameters in traditional methods, making train tracking control more efficient and safer. However, this article lacks research on the robustness of the train tracking process, which can be studied in the author's future research.

Author Contributions: Conceptualization, resources, methodology, validation, formal analysis, investigation, data curation, writing—original draft preparation and visualization, are provided by B.C.; writing—review and editing are provided by L.Z. and J.C.; software and data curation are provided by B.C., G.C. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62003150).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors are greatly thankful to Chunhai Gao for the project administration and funding acquisition. Also, the authors are greatly thankful to the reviewers and editor for their precious advice to improve the quality of the study.

Conflicts of Interest: Authors Gaoyun Cheng and Yiqing Liu are employed by the company Traffic Control Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PID	Proportional Integral Derivative
MPC	Model Predictive Control
SAC	Soft Actor-Critic
RL	Reinforcement Learning
ATC	Automatic Train Control
DDPG	Deep Deterministic Policy Gradient

Appendix A

The process of the SAC algorithm is shown as follows:

Algorithm A1: SAC algorithm

1. Initialization: $iter = 1$, Q-function parameters θ_1 and θ_2, policy weights ϕ, replay buffer $\mathcal{D} = \emptyset$ and target network weights $\bar{\theta}_1 = \theta_1$, $\bar{\theta}_2 = \theta_2$

while $iter \leq MAX_{iter}$ **do**

2. $s_t = s_0$, $timestep = 1$

while $timestep \leq MAX_{timestep}$ **do**

3. sample a_t from policy π_ϕ, $a_t \sim \pi_\phi(a_t \mid s_t)$
4. obtain r_t and s_{t+1} by inputting a_t into the environment
5. store (s_t, a_t, r_t, s_{t+1}) into $\mathcal{D} = [(s_0, a_0, r_0, s_1), (s_1, a_1, r_1, s_2), ...]$
6. $s_t = s_{t+1}$
7. $timestep = timestep + 1$

end

8. $updactor = 1$

while $updactor \leq MAX_{updactor}$ **do**

9. update the Q-function parameters θ_1 and θ_2 by Equation (6)
10. update policy weights ϕ by Equation (7)
11. update temperature parameter α by Equation (8)
12. update target network weights $\bar{\theta}_1$ and $\bar{\theta}_2$ by $\bar{\theta}_i = \tau\theta_i + (1 - \tau)\bar{\theta}_i$, for $i \in 1, 2$
13. $updactor = updactor + 1$

end

end

14. Output θ_1, θ_2, ϕ

References

1. Bock, U.; Varchmin, J.-U. Improvement of line capacity by using "virtually coupled train formations". *VDI Berichte* **1999**, *1488*, 315–324.
2. Bock, U.; Bikker, G. Design and development of a future freight train concept—"Virtually coupled train formations". *IFAC Proc. Vol.* **2000**, *33*, 395–400. [CrossRef]
3. König, S.; Bikker, G. Developing and Implementing a Framework for CASE Tool Coupling-Object Orientation upon Tool Level. In Proceedings of the European Concurrent Engineering Conference, Leicester, UK, 17–19 April 2000.
4. Bock, U.; Varchmin, J.U. Virtually coupled train formations: Wireless communication between train units. In Proceedings of the General Traffic Forum, Braunschweig, Germany, 6–7 April 2000.
5. Bikker, G.; Bock, U. Einsatz eines Prozeßmodells zur Analyse und Spezifikation von Bussystemen. *EKA* **1999**, *99*, 509–526.
6. Cao, Y.; Wen, J.; Ma, L. Tracking and collision avoidance of virtual coupling train control system. *Future Gener. Comput. Syst.* **2021**, *120*, 76–90. [CrossRef]
7. Cao, Y.; Yang, Y.; Ma, L.; Wen, J. Research on Virtual Coupled Train Control Method Based on GPC & VAPF. *Chin. J. Electron.* **2022**, *31*, 897–905.
8. Lin, P.; Huang, Y.; Zhang, Q.; Yuan, Z. Distributed velocity and input constrained tracking control of high-speed train systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *51*, 7882–7888. [CrossRef]
9. Liu, Y.; Zhou, Y.; Su, S.; Xun, J.; Tang, T. An analytical optimal control approach for virtually coupled high-speed trains with local and string stability. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 102886. [CrossRef]
10. Luo, X.; Tang, T.; Liu, H.; Zhang, L.; Li, K. An Adaptive Model Predictive Control System for Virtual Coupling in Metros. *Actuators* **2021**, *10*, 178. [CrossRef]
11. Wang, Q.; Chai, M.; Liu, H.; Tang, T. Optimized Control of Virtual Coupling at Junctions: A Cooperative Game-Based Approach. *Actuators* **2021**, *10*, 207. [CrossRef]
12. Chen, Y.; Huang, D.; Li, Y.; Feng, X. A novel iterative learning approach for tracking control of high-speed trains subject to unknown time-varying delay. *IEEE Trans. Autom. Sci. Eng.* **2020**, *19*, 113–121. [CrossRef]
13. Felez, J.; Vaquero-Serrano, M.A.; de Dios Sanz, J. A Robust Model Predictive Control for Virtual Coupling in Train Sets. *Actuators* **2022**, *11*, 372. [CrossRef]
14. Su, S.; She, J.; Li, K.; Wang, X.; Zhou, Y. A nonlinear safety equilibrium spacing-based model predictive control for virtually coupled train set over gradient terrains. *IEEE Trans. Transp. Electrif.* **2021**, *8*, 2810–2824. [CrossRef]

15. Chengwei, W.; Weiran, Y.; Wensheng, L.; Wei, P.; Guanghui, S.; Xie, H.; Wu, L. A Secure Robot Learning Framework for Cyber Attack Scheduling and Countermeasure. *IEEE Trans. Robot.* **2023**, *39*, 3722–3738.
16. Wu, C.; Pan, W.; Staa, R.; Liu, J.; Sun, G.; Wu, L. Deep reinforcement learning control approach to mitigating actuator attacks. *Automatica* **2023**, *152*, 110999. [CrossRef]
17. He, Y.; Lv, J.; Liu, H.; Tang, T. Toward the Trajectory Predictor for Automatic Train Operation System Using CNN–LSTM Network. *Actuators* **2022**, *11*, 247. [CrossRef]
18. He, Y.; Lv, J.; Tang, T. Communication-Based Train Control with Dynamic Headway Based on Trajectory Prediction. *Actuators* **2022**, *11*, 237. [CrossRef]
19. Huang, Z.; Wang, P.; Zhou, F.; Liu, W.; Peng, J. Cooperative tracking control of the multiple-high-speed trains system using a tunable artificial potential function. *J. Adv. Transp.* **2022**, *2022*, 3639586. [CrossRef]
20. Li, Z.; Yin, C.; Ji, H.; Hou, Z. Constrained spatial adaptive iterative learning control for trajectory tracking of high speed train. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 11720–11728. [CrossRef]
21. Zhou, Y.; Zhang, J.; Yang, H. Research on Tracking Control of Urban Rail Trains Based on Improved Disturbance Observer. *Appl. Sci.* **2023**, *13*, 7403. [CrossRef]
22. Wang, X.; Li, S.; Cao, Y.; Xin, T.; Yang, L. Dynamic speed trajectory generation and tracking control for autonomous driving of intelligent high-speed trains combining with deep learning and backstepping control methods. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105230. [CrossRef]
23. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
24. Chen, B.; Gao, C.; Zhang, L.; Chen, J.; Chen, J.; Li, Y. Optimal Control Algorithm for Subway Train Operation by Proximal Policy Optimization. *Appl. Sci.* **2023**, *13*, 7456. [CrossRef]
25. Li, Y.; Ang, K.H.; Chong, G.C. PID control system analysis and design. *IEEE Control. Syst. Mag.* **2006**, *26*, 32–41.