agronomy

MDPI

Article

Supervised Reinforcement Learning-Based Collaborative Master–Slave Harvest Control Study in Wheat

Zhikai Ma [1,2], Chao Zhang [2], Wei Wang [2], Hao Wang [3], Helong Yu [1] and Chunjiang Zhao [1,3,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] College of Mechanical and Electrical Engineering, Agricultural University of Hebei, Baoding 071001, China
[3] Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
* Correspondence: zhaocj@nercita.org.cn

Abstract: Aiming at the difficulty of controlling the longitudinal relative position of agricultural machines during the agricultural master–slave navigation cooperative operation and the weak adaptability of the unitary traditional control method in the face of the working conditions of complex farmland environments, this paper proposes a supervised reinforcement learning (SRL)-based longitudinal stable and safe control method applicable to master–slave navigation harvesting and unloading operations. Firstly, to improve the algorithm training success rate, a supervisor trained on actual driving data is introduced into the actor–critic reinforcement learning method. Secondly, in order to improve the effect of agricultural machine operation, considering the actual grain unloading operation scene and combining the smoothness of operation and the safety of unloading, a new reward function in the supervised reinforcement learning algorithm is designed. Finally, the performance of the proposed SRL control strategy is verified by simulation and agricultural machines following grain unloading tests. The results of field operation show that, when the harvester speed is 1.2 m/s, the average absolute deviation of the actual distance between the two trucks is 0.048 m, and the maximum deviation of the steady state is 0.26 m. In the variable speed test, when the harvester speed is 0.4 m/s and 1.2 m/s, the average absolute deviation of the actual distance between the two trucks is 0.079 m and 0.091 m, and the maximum deviation of the steady state is 0.20 m and 0.21 m, and the cooperative accuracy can fulfill the operational demands of harvesting cooperative unloading. The study's results can serve as a technological reference for autonomous harvesting operations in the field.

Keywords: following control; automatic guidance; longitudinal stability control; safe unloading of grain

check for updates

Citation: Ma, Z.; Zhang, C.; Wang, W.; Wang, H.; Yu, H.; Zhao, C. Supervised Reinforcement Learning-Based Collaborative Master–Slave Harvest Control Study in Wheat. *Agronomy* **2024**, *14*, 2577. https://doi.org/10.3390/agronomy14112577

Academic Editor: Gniewko Niedbała

Received: 20 September 2024
Revised: 17 October 2024
Accepted: 30 October 2024
Published: 1 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With rising agricultural labor expenses and an aging population, labor shortages have hampered agricultural growth. Intelligent agricultural equipment, especially unmanned and less-manned operation equipment, can greatly improve labor efficiency, solve the labor shortage problem, improve agriculture productivity, protect the environment, and safeguard national food security. It is an important part of intelligent agriculture and is essential to the future development of agriculture. Therefore, researching intelligent, automated, and efficient agricultural equipment is an urgent response to contemporary agriculture and a trend that cannot be avoided [1–4]. As modern agriculture advances, the requirement for intelligent farming machinery is increasing; nevertheless, current intelligent equipment is relatively underdeveloped, necessitating urgent study into the intelligence of agricultural technology [5–8].

In the past few years, researchers have looked into intelligent control technology for farm equipment and automatic grain unloading technology [9,10], as well as satellite,

LiDAR, and visual positioning and navigation methods, which have led to good control results. However, most studies on intelligent control primarily concentrate on the autonomous operation of individual machines, with comparatively fewer studies exploring the cooperative control of multiple machines. Cooperative operation of several farm machinery, whether of the same or different species, in the field can minimize the waiting and idle time of farm machines, greatly enhance their efficiency, and ultimately decrease the need for human labor. Therefore, multiple farm machines of the same or different species of agricultural machinery working together are considered an effective solution to meet the above new agricultural demands.

In 2000, Kyoto University in Japan created an automated tracking system for agricultural machinery [11], in which the driver operates one agricultural machine to travel ahead, and an onboard controller controls the other to complete the following operation. Multi-machine cooperation has been extensively researched in both academia and industry. The critical technologies of multi-machine cooperative operation include cooperative mode of operation, operation path planning [12], communication technology [13], operation task allocation [14], cooperative control, and remote monitoring platform [15], etc. Among them, the collaborative operation modes can be divided into two types: free mode and master–slave follower mode [16,17]. Compared with the free mode, the master–slave–follower mode is more widely applicable, such as harvesters–grain trucks, seeding trucks–seed fertilizer replenishment trucks, etc. Multi-machine cooperative control technology is the key and a difficult problem that researchers have to solve. National and international scholars have undertaken much exploration and research on cooperative control algorithms.

Lee at the University of Florida used a Proportional-Integral-Differential (PID) algorithm to control the synchronized steering of two vehicles and a system of vision to achieve forward motion of the two vehicles [18]. The experimental results demonstrate that the control system enables two test vehicles following each other to advance at varying speeds with an error range of 11 to 32 cm. Zhang et al. at Hokkaido University designed a leader-follower system based on two agricultural robots, facilitating the collaboration of the two robots to execute tasks such as tracking in a linear trajectory and inverting into the ground, while guaranteeing the absence of any collisions. [19]. Luo at the University of Ottawa investigated a master–slave vehicle corner control law and a velocity control law for the slave vehicle. This allowed two robots to complete their respective path-tracking tasks with offsets of less than 0.1 m [20]. Li from China Agricultural University, designed an automatic navigation system for agricultural machines consisting of a master and a slave [21], and the controller calculates the lateral and heading deviations of both based on the position information of the master and the slave. The experimental results show that the root-mean-square error is 6.76 cm when the slave vehicle follows the master vehicle at a speed of 0.8 m/s. The system has a stable performance and meets the field operation requirements. The work conducted by Bai Xiaoping and his team from Shenyang Institute of Automation, Chinese Academy of Sciences [22,23], involved a systematic investigation of the combined harvester pilot-following structure. They integrated the feedback linearization theory and the sliding mode control theory to develop the control law for harvester fleet formation keeping and path tracking. This design enabled the agricultural machine to operate with high precision. Xu Guangfei [24] developed a tractor master–slave follower controller based on a model predictive control algorithm, which can effectively respond to factors such as environmental uncertainty and disruption on the basis of securing the master–slave follower of an agricultural machine. However, the vehicle modelling is complicated, and there is still the defect of untimely response in the farmland. Jun Chen [25] of Northwest Agriculture and Forestry University applied neural network control technology to the automatic driving of agricultural vehicles, which improved the controller's ability to adapt to the environment but did not consider the neural network's powerful learning ability of the driver's operating habits.

The studies mentioned above still have some limitations. Firstly, part of the research focuses on traditional controllers such as PID, sliding mode variable structure, and model

predictive control, and the traditional control algorithms are ineffective and fails to correctly regulate the relative distance between vehicles. Secondly, although the back-propagation (BP) neural network-based controller [26] has a fast response and high environmental adaptability, it uses a simple neural network structure. It needs to use the powerful learning ability of neural networks fully. Currently, reinforcement learning (RL) has solved many academic and industrial challenges and achieved specific research results under the excellent feature representation capability of neural networks [27,28], which can provide quasi-optimal control performance in changing environments by use of trial-and-error processes. However, the slow convergence of RL is an issue that needs to be addressed and may lead to some inefficient situations. Supervised reinforcement learning that introduces supervised learning can alleviate the problem of slow RL convergence and shorten the learning time by incorporating the advantages of supervised learning (SL). The supervisor introduced in the actor–critic network guides the network's learning process, evaluates the network's performance, and helps the network avoid problems such as overfitting, which improves the network's generalization ability and prediction accuracy, enables more complex learning tasks, and improves system performance. This paper employs a master–slave cooperative control algorithm grounded in supervised reinforcement learning technology to maximize the learning capabilities of neural networks, thoroughly explore and enhance their potential, and ensure safe and stable master–slave cooperative harvesting and unloading of grain. The supervisor in the actor–critic networks directs the learning process, assesses network performance, and mitigates issues like overfitting, thereby enhancing the network's generalization capability and predictive accuracy, facilitating more intricate learning tasks, and boosting overall system performance.

The main contributions of this paper are: The SRL method was explored and applied to the speed control of real agricultural machines, and a supervised reinforcement learning-based control strategy for collaborative multi-vehicle grain unloading was developed. In order to achieve smooth and safe loading of grain in the grain truck, combining the smoothness of operation and the safety of unloading, a new reward function in the SRL method is designed to deal with the problem of weak adaptability of working conditions in complex farmland operating environments. With the introduction of supervisors, reinforcement learning algorithms converge faster, and the control response is better than that of traditional linear control algorithms.

2. Models and Methods

2.1. Master–Slave Unloading Control

The cooperative longitudinal control strategy adopts a layered structure, where the upper controller generates the desired acceleration of the grain truck derived from the relative speeds of the harvester and the grain truck, the relative acceleration of the harvester and the grain truck, the error in the workshop distance between the harvester and the grain truck, and the error between the actual speed and the expected speed of the grain truck, and the lower controller achieves the desired acceleration through the coordinated control of the valve openings and the braking pressure [29,30]. In the actual following control, the grain truck throttle opening and braking pressure corresponding to the grain truck's desired acceleration are obtained by querying a pre-calibrated table of the correspondence between acceleration and throttle opening to achieve the goal of the grain truck following the harvester stably. The master–slave cooperative control framework of harvesters and grain carriers is shown in Figure 1 below.

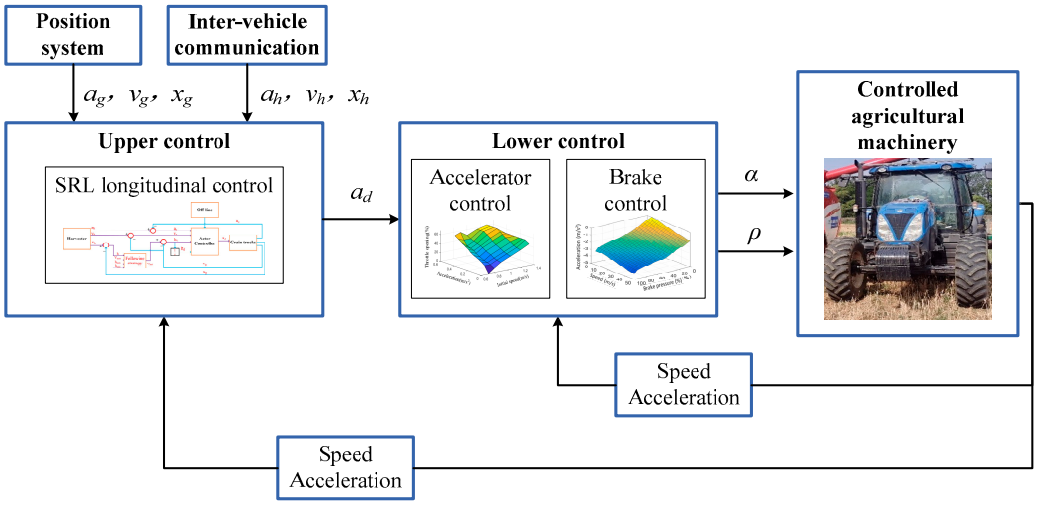


Figure 1. Longitudinal control framework of SRL-based master–slave cooperative system. Note: a_g, v_g, x_g are the acceleration, velocity, and position of the grain truck; a_h, v_h, x_h are the acceleration, velocity, and position of the harvester truck; a_d is the desired acceleration of the grain truck; α and ρ are throttle opening and brake pressure information.

Master–Slave Collaborative Control Framework of Harvester and Grain Truck

The control flow of the SRL longitudinal controller is shown in Figure 2 below [31]. The positioning system collects the two machines' acceleration, velocity, and position information, and the inter-vehicle communication equipment sends the information collected by the positioning system on the harvester to the grain truck controller. The longitudinal following grain truck controller in Figure 2 uses an actor network trained offline by a supervised reinforcement learning network. The desired acceleration issued by the actor network is looked up in the table to obtain the actual acceleration of the following tractor as it passes through the acceleration following the controller. Finally, the corresponding throttle opening of the tractor is found, thus controlling the speed of the grain truck.

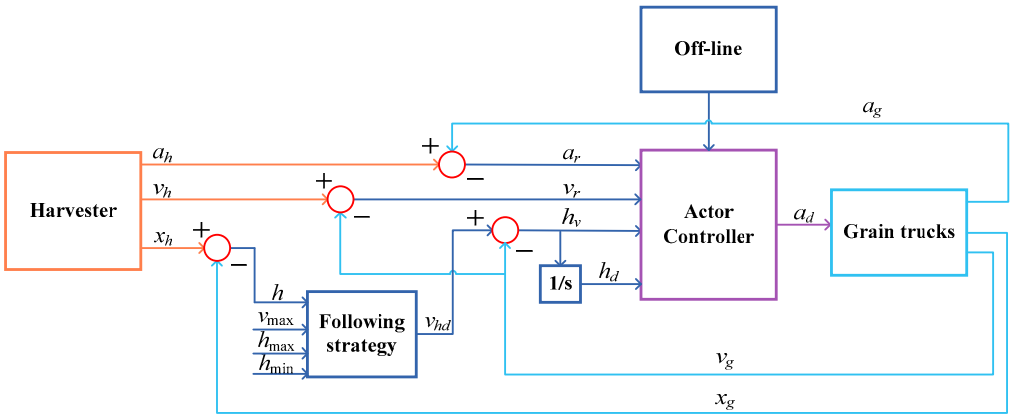


Figure 2. SRL longitudinal control flow graph. Note: h_{max} is the maximum distance between the two vehicles, h_{min} is the minimum distance, h is the actual distance; v_{max} is the maximum speed of the grain truck, v_{hd} is the desired speed of the grain truck; v_r, a_r, h_d, and h_v are the relative speeds of the harvester and the grain truck, the relative acceleration of the harvester and grain truck, the workshop distance error between the harvester and the grain truck, and the error between the desired speed and the actual speed of the grain truck, respectively.

2.2. Supervised Reinforcement Learning

Machine learning is a multi-disciplinary cross-discipline and an essential sub-field of artificial intelligence. As with reinforcement learning, supervised learning is a widely

employed technique in machine learning that finds extensive applications in addressing classification and regression issues [32,33]. Supervised learning can use labeled training data to find mapping relationships between input and output functions. Through a trial-and-error mechanism, reinforcement learning can obtain quasi-optimal control performance in changing environments. By introducing supervised learning into reinforcement learning to form supervised reinforcement learning, it can give full use to the advantages of both, make up for the shortcomings of each, improve the learning efficiency, and make the system more intelligent and robust.

2.2.1. Supervised Reinforcement Learning Framework

The supervised reinforcement learning framework is shown in Figure 3, which consists of four modules: actor, critic, supervisor, and environment [34]. Among them, the actor network is responsible for establishing the mapping link between the system state and the control signal. This mapping enables the control action to be outputted based on the current prevailing system state. The controlled system responds to this state and issues a new action. The critic network evaluates the control signals from the network and the reward and punishment signals from the system to update the actor's control strategy. In one hand, the supervisor forms a composite action with the actor's output action and then applies it to the controlled system to ensure the minimum vehicle traveling safety and speed up the network training. On the other hand, the supervisor, in conjunction with the current state of the system, provides advice to the actor on which actions are advantageous to take to reduce the actor's exploratory nature in the trial-and-error phase. The environment corresponds to the controlled system, which in this paper refers to grain truck.

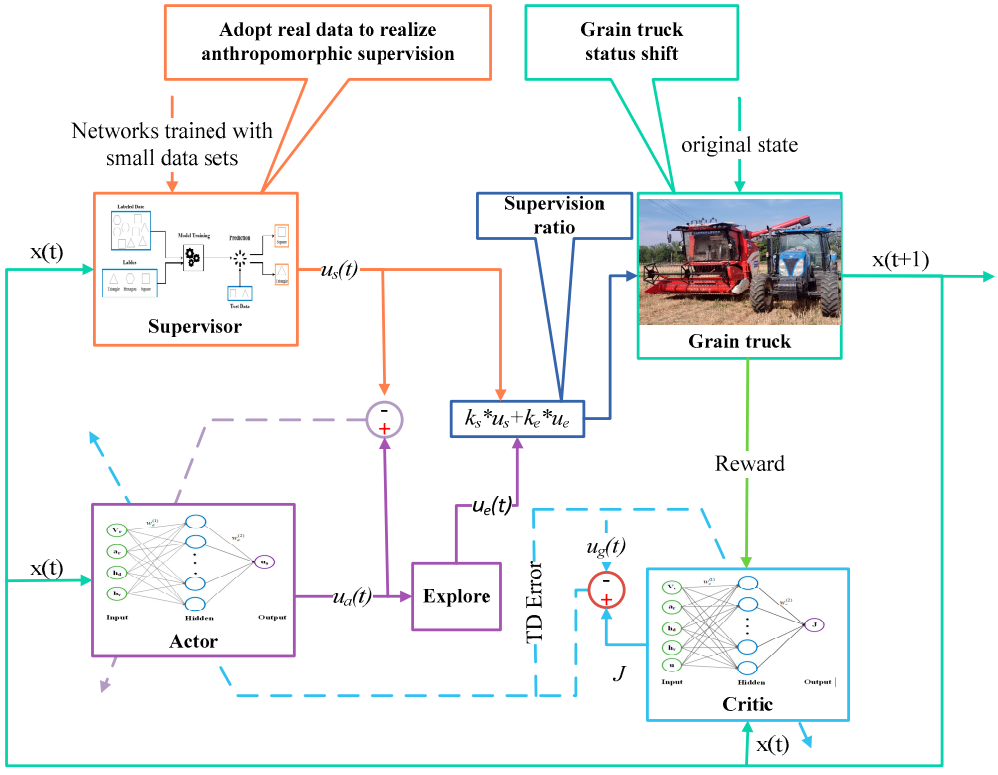


Figure 3. Schematic diagram of the SRL framework.

2.2.2. Actor Neural Network and Critic Neural Network

Both the actor network and the critic network use a three-layer feed forward neural network. The activation function for the hidden layer nodes is a hyperbolic tangent function, and the weights are updated using error back-propagation. The input state quantities of the actor network and the critic network are x(t) = [v_r, a_r, h_d, h_v] and x(t) = [v_r, a_r, h_d, h_v, u],

respectively. Their output state quantities are u_a and J. The structure of the actor network and critic network is seen in Figure 4, and the specific formulas are referred to in the literature [35].

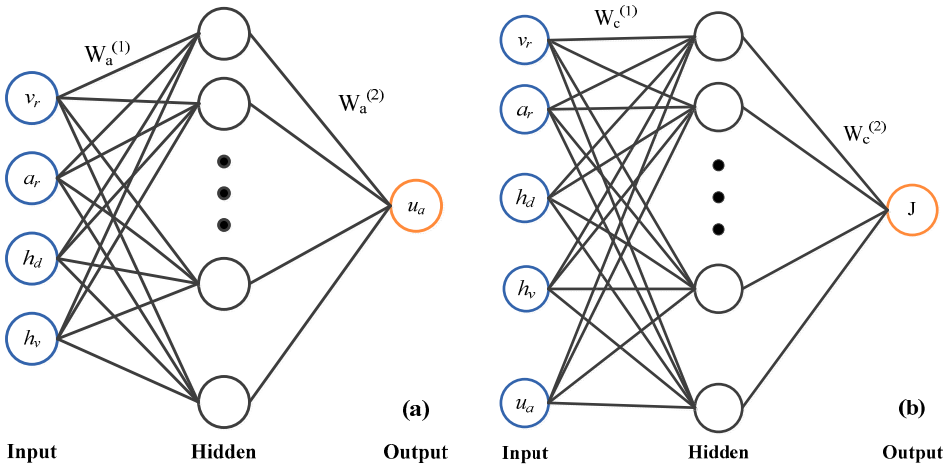


Figure 4. (**a**) Actor neural network; (**b**) Critic neural network.

The reward function is an integral objective optimization function in reinforcement learning, used to represent which actions are better from a long time perspective, as follows:

$$\mathrm{R(t)} = \mathrm{r(t+1)} + \gamma r(t+2) + \gamma r(t+3) + \ldots = \sum_{k=0}^{T} \gamma^k r(t+k+1) \quad (1)$$

Discount factor γ is used to determine the relationship between immediate and long-term rewards.

Under the ideal stable following situation, the error between the desired speed and the actual speed of the grain truck should be equal to 0. After combining the parameters of smoothness of operation and safety of unloading, the objective reward function is designed as follows:

$$\mathrm{r(t)} = -(k_1 h_d^2 + k_2 h_v^2 + k_3 a_r^2 + k_4 F_x^2 + k_5 F_d^2) \quad (2)$$

where k_1, k_2, k_3, k_4, k_5 are positive coefficients, which are taken according to the actual training. F_x is the stability function and F_d is the security function, as explained in later sections.

2.2.3. Supervisor Network

Supervised reinforcement learning is the integration of the benefits of reinforcement algorithmics with supervised learning. The actor network learns from interaction with the environment and from feedback from supervisors. In supervised reinforcement learning, there are three ways in which the supervisor acts on the system network: (1) Shaping: Supervisors give extra rewards, thus speeding up the process of learning in the actor network. (2) Nominal control: The supervisor skips the actor network and applies control signals directly to the controlled system. (3) Exploration: The supervisor gives the actor network cues to choose the action to take.

The training set samples of the supervisor network are [v_r, a_r, h_d, h_v, a_d]; a_d is the desired acceleration, and the training data are collected from the operation data of real drivers [36]. The training set data were trained using MATLAB's neural network fitting; the number of neurons in the hidden layer of the neural network was the default value of 10, and the training algorithm was the Levenberg–Marquardt algorithm, with the maximum number of iterative steps and the performance target set to 1000 and 0.001, respectively, at training time.

2.2.4. Composite Control

The control signal of the controlled system is derived from the weighted sum of the actor network and the output control quantities with the following formula:

$$u(t) = k_s u_s(t) + k_e u_e(t) \tag{3}$$

where u(t) is the composite control quantity, $u_e(t)$ is the explored actor output control quantity, $u_e(t) = u_a(t) + N(\mu,\sigma)$, mean $\mu = 0$, variance σ is the random variable, $u_s(t)$ is the supervisor output control quantity. k_s and k_e are the weight coefficient, $k_s + k_e = 1$, taking the value range [0, 1].

2.2.5. Weight Values Updating

The weights of the actor network and the critic network during training are updated by error back-propagation, and the weight update error of the actor has two parts: the evaluation error rewarded by the critic in reinforcement learning and the supervisory error formed by the combined control of the actor network and the supervisor. The weights of the actor are updated according to the following rule as [37]:

$$\omega_a(t+1) = \omega_a(t) + k_s\, \omega_a^{RL}(t) + k_e \Delta\omega_a^{SL}(t) \tag{4}$$

in which $\omega_a^{RL}(t)$ and $\omega_a^{SL}(t)$ are the updates based on RL and the supervised learning, respectively.

The error and objective function of RL are defined as:

$$e_a^{RL}(t) = J(t) - u_g(t) \tag{5}$$

$$E_a^{RL}(t) = \frac{1}{2}\left(e_a^{RL}(t)\right)^2 \tag{6}$$

in which u_g(t) is the desired control objective. Here, u_g(t) is set to 0, because the reward tends to zero if optimal actions are being taken.

The supervisory error and its objective function are calculated as:

$$e_a^{SL}(t) = u_s(t) - u_a(t) \tag{7}$$

$$E_a^{SL}(t) = \frac{1}{2}\left(e_a^{SL}(t)\right)^2 \tag{8}$$

The RL and supervisory objective functions are minimized by the gradient descent approach, and thus the reinforcement-based update and the supervisory update are calculated as:

$$\Delta\omega_a^{RL}(t) = l_a(t)\left(-\frac{\partial E_a^{RL}(t)}{\partial \omega_a(t)}\right) = -l_a(t)\frac{\partial E_a^{RL}(t)}{\partial J(t)}\frac{\partial J(t)}{\partial u(t)}\frac{\partial u(t)}{\partial \omega_a(t)} \tag{9}$$

$$\Delta\omega_a^{SL}(t) = l_a(t)\left(-\frac{\partial E_a^{SL}(t)}{\partial \omega_a(t)}\right) = -l_a(t)\frac{\partial E_a^{SL}(t)}{\partial u_a(t)}\frac{\partial u_a(t)}{\partial \omega_a(t)} \tag{10}$$

in which l_a(t) is the learning rate of the actor network at time t, $0 < l(t) < 1$.

The critic network is to update the prediction error by defining the timing difference error as a weight. For the critic network, the prediction error has the same expression with the temporal difference (TD) error as follows:

$$ec(t) = \beta J(t) - (J(t-1) - r(t)) \tag{11}$$

in which β is the discount factor, $0 < \beta < 1$.

The squared error objective function is calculated as

$$Ec(t) = \frac{1}{2}e_c^2(t) \tag{12}$$

$$\Delta\omega_c(t) = l_c(t)\left(-\frac{\partial E_c(t)}{\partial\omega_c(t)}\right) = -l_c(t)\frac{\partial E_c(t)}{\partial J(t)}\frac{\partial J(t)}{\partial\omega_c(t)} \tag{13}$$

in which l_c(t) is the learning rate of the critic network at time t, $0 < l_c(t) < 1$.

2.3. Follower Strategy and Reward Function

2.3.1. Grain Truck Follower Strategy

Unmanned master–slave cooperative wheat harvesting involves fixed unloading ports that can lead to a concentrated grain accumulation. After the grain is piled up to a certain extent, wheat will overflow in places close to the unloading opening. Empty bins will appear in places far away from the unloading opening, thus leading to wheat being piled up unevenly in the grain truck and reducing the volume utilization rate [38].

For the above reasons, this paper adopts the distance-speed model of safe vehicle spacing. The system regulates the speed of the harvester to modify the desired distance between the two vehicles, thus realizing the goal of reciprocating motion and uniform unloading of grain at the unloading port within the preset unloading range, with the following equation:

$$v(h) = \begin{cases} 0 & , \quad h < h_{\min} \\ v_{\max}\frac{h-h_{\min}}{h_{\max}-h_{\min}} & , \quad h_{\min} < h < h_{\max} \\ v_{\max} & , \quad h > h_{\max} \end{cases} \tag{14}$$

where $h_{\max}$ is the maximum distance between the two vehicles, that is, the length of the grain truck box plus the minimum distance; $h_{\min}$ is the minimum distance, that is, the distance between the positioning system of the harvester and the positioning system of the grain truck; h is the actual distance; $v_{\max}$ is the maximum speed of the grain truck. Positioning system installation location: the harvester positioning system is installed at the unloading port, and the grain truck positioning system is installed at the front of the tractor.

2.3.2. Transfer of State Equation

When the grain truck is in the x(t) state, the relationship between the position, speed, and acceleration of the truck is as follows [39]:

$$\begin{cases} \dot{x}_g(t) = v_g(t) \\ \dot{v}_g(t) = a_g(t) \end{cases} \tag{15}$$

The relational equation for the change in the speed of the grain truck from the x(t) state to the x(t + 1) state is as follows:

$$v_g(t+1) = v_g(t) + a_g(t)\Delta t \tag{16}$$

Consider that when the grain truck follows the driver, the controller of the grain truck can not only measure its acceleration, the relative speed of the two trucks, and the workshop distance through the positioning system but also obtain the acceleration and speed information of the harvester through the workshop communication equipment. The state quantity of the following system model is selected as x(t) = [v_r, a_r, h_d, h_v].

The next moment state quantities are as follows:

$$\begin{cases} v_r(t+1) = v_h(t+1) - v_g(t+1) \\ a_r(t+1) = a_h(t+1) - a_g(t+1) \\ h_v(t+1) = v(t+1) - v_g(t+1) \\ h_d(t+1) = h_d(t) + h_v(t+1)\Delta t \end{cases} \tag{17}$$

where v is the desired speed of the grain truck, and Δt is the sampling period.

2.3.3. Stability

The size of the grain truck's acceleration and deceleration have a significant impact on the smoothness of grain unloading. If the grain truck can move gently with acceleration with a small absolute value when picking up grain, the grain unloading will be smoother. Corresponding to the fluctuating characteristics of the smaller values is the trend toward more minor acceleration changes. Define the operating stability function as follows [40]:

$$F_d = \frac{d^2}{400} \tag{18}$$

The square of the fluctuation value is divided by a base value (400), thus scaling the feature to the range [0–1]. The base value was determined in the following way: the data sampling interval was 0.2 s. The acceleration range was between -2 and 2 m/s^2 because the speeds of the agricultural machines were generally low during their field operations; the maximum fluctuation value, d, was defined as the ratio of the acceleration range length to the data sampling time, i.e., $d = 2 - (-2)/0.2 = 20$ m/s^3, and the maximum fluctuation squared to obtain the base value of 400.

2.3.4. Grain Unloading Safety

When the discharging port of the harvester is in close proximity to the periphery of the grain carrier box, due to the constant shaking of the grain carriage itself in the field, it is easy for the unloading port to go outside the carriage and thus unload the grain outside the carriage as well. In order to minimize the occurrence of the above so that the unloading opening of the harvester is always within the length of the carriage of the grain truck, the unloading safety equation is defined as follows:

$$F_X = \begin{cases} 0 & , \quad h_{\min} + h_y \le h \le \mathrm{L} + h_{\min} - h_y \\ -1 & , \quad \text{otherwise} \end{cases} \tag{19}$$

where L is the length of the grain truck box and h_y is the unloading safety threshold.

2.4. Experiment Scenarios and Data Acquisition

To verify the validity and adaptability of the established model, a real-vehicle test platform is constructed by combining the independently modified harvester and grain trucks; the technical parameters of the harvester and grain truck are shown in Table 1. The actual vehicle test was conducted on 10 June 2024 in the test field of Malan Farm, Xinji City, Hebei Province, which has a flat and open field surface with no shelter, as shown in Figure 5. The traveling data of the two vehicles collected by the positioning system is converted into CAN data by the controller and sent to the software of the host computer and saved, after which the collected data are filtered and processed, and after removing the noise, data collection anomalies, and other interferences, the following speed-following diagrams between the harvester and the grain truck under different speed conditions are obtained.

Table 1. Technical parameters of the harvester and the grain truck.

Experimental Trucks	Parameter	Technical Specification
Harvester	Model	4LZ-10M7(G4)
	Dimensions (mm × mm × mm)	7000 × 3385 × 3470
	Overall weight (kg)	7250
	Engine power (ps)	200
	Cutting width (mm)	3000
	Feeding quantity (kg/s)	10
	Granary volume (m^3)	2.8
	Driving wheelbase(mm)	1900

Table 1. *Cont.*

Experimental Trucks	Parameter	Technical Specification
Grain truck	Model	T1404
	Overall dimensions (mm × mm × mm)	5290 × 2414 × 3115
	Maximum usable mass (kg)	10,000
	Minimum usable mass (kg)	5848
	Engine power (ps)	142–178
	Granary volume (m^3)	3.1
	Wheelbases (mm)	2739



Figure 5. Real-vehicle test on farmland.

3. Results

3.1. Simulation Results and Analysis

3.1.1. Training Process

The supervising reinforcement learning model with an initial learning rate of 0.3 for both the actor network and critic network, decreased by 0.05 per time step and finally stopped at 0.001. The discount factor is selected to 0.9, and the training is carried out by constructing a typical working condition of a harvester, as shown in Figure 6. During the training process, the harvester traveled according to the training conditions, and the grain truck trailed the harvester, which had its initial speed set at 0.6 m/s. The harvester is accelerating at 0.1 m/s^2 for 9 s in the 3rd s, accelerating to 1.5 m/s, traveling at 1.5 m/s for 5 s, and decelerating at -0.2 m/s^2 for 2 s, decelerating to 1.1 m/s. The harvester travels at 1.1 m/s for 5 s and decelerates to 0.6 m/s for 2 s with an acceleration of -0.25 m/s^2. The harvester then experiences one sine acceleration fluctuation and travels maintaining a consistent velocity of 0.6 m/s. Although the simulation training is conducted in simple settings, the training process is intensive and, via trial and error, will drive the actor to go through many distinct states, making the trained controller broadly versatile.

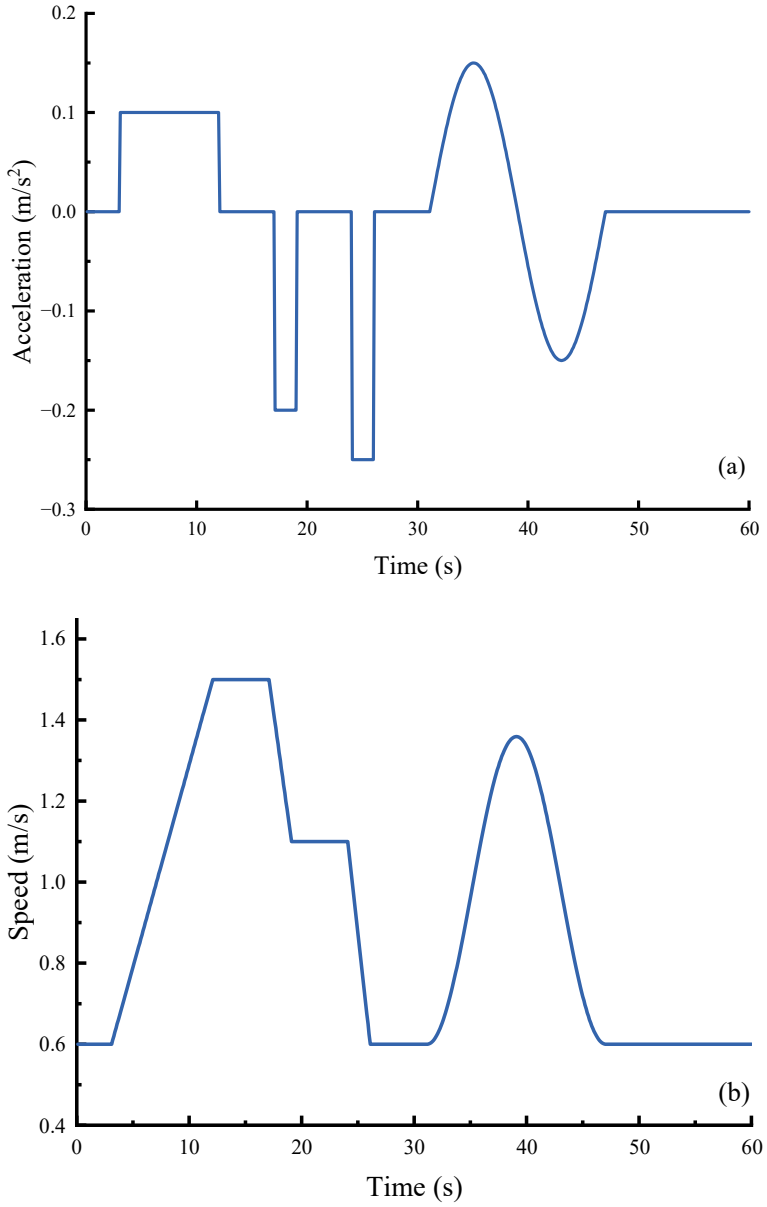


Figure 6. Diagram of training conditions. (**a**) Acceleration; (**b**) Speed.

3.1.2. Comparison of Supervised Reinforcement Learning and Reinforcement Learning

Verification of the convergence of the SRL algorithm following the implementation of the supervisor is conducted by simulated training comparative experiments on SRL and RL with the same aim. In order to give the actor more time to learn, the maximum number of training steps for an experiment is 1000. The training is successful when the actor network and the critic network weights start to remain fixed, the learning rate decreases to 0.03, the value of the reward function continues to stabilize under 0.001, and the grain truck can follow the harvester under the standard working conditions, the speed error converges within 0.05 m/s and the number of training times $\leq$ 1000. Table 2 demonstrates the performance comparison between SRL and RL; the presence of a supervisor during training allows for stable convergence.

Table 2. Comparison of training results for supervised reinforcement learning (SRL) and reinforcement learning (RL).

Type	Training Time	Training Success Rate
SRL	100	100%
RL	100	60%

3.1.3. Evidences for the Characterization of the Reward Function

In order to verify the difference between the new reward function, which combines smoothness of operation and safety of unloading, and the old reward function, a simulation training comparison experiment was carried out between the new reward function and the old reward function. Simulation sets the maximum speed of the vehicle at 1.5 m/s, the acceleration range [−1.5, 1.5], the maximum distance h_{max} is 4.5, the minimum distance h_{min} is 3.5 m, the safety threshold is 0.1 m, and the length of the grain truck compartment is 1 m. When the new reward function parameters k_4 and k_5 take the value of 0, it is the old reward function.

Observing the distance adjustment diagram, it can be concluded that after considering the safety of grain unloading, the new reward function realizes the orderly adjustment of the grain wagon compartment relative to the unloading port of the harvester wagon and the maximum actual distance between the wagons is changed from 4.5 m to 4.4 m, which leaves enough safe distance for unloading, and effectively reduces the situation of unloading grain into the outside of the compartment when unloading. The speed of the two vehicles in Figure 7 reaches a maximum speed of 1.5 m/s, at which time the maximum distance between the vehicles is 4.5 m. It can be observed from the figure that the maximum distance between the vehicles under the old reward function is 4.5 m. In contrast, the new reward function, which considers the safety thresholds, only reaches a maximum distance of 4.4 m at a speed of 1.5 m/s, which aligns with the desired design objective.

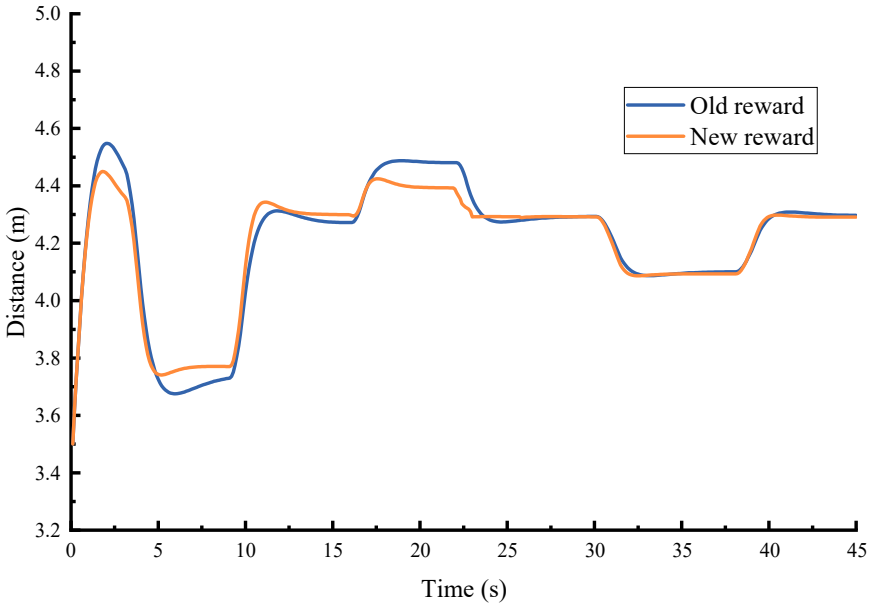


Figure 7. Diagram of the distance adjustment region.

To assess the impact of the smoothness parameter in the innovative reward function on the model, the two models with the two previously mentioned reward functions will be applied respectively for simulation comparison for the same set of control quantities and the speed trajectory of the harvester vehicle for many times, and the results are shown in Figure 8. Observing Figure 8a, it can be concluded that after considering the smoothness of the operation, the grain truck tends to accelerate and decelerate with a smaller absolute

value of acceleration to complete the unloading operation. From Figure 8b, it can be seen that the probability distribution of the absolute value of acceleration is more significant than 0.1 for the old reward function without considering that the unloading safety is more significant than that of the new reward function and the acceleration value is concentrated in the region near zero. It means that the old reward function tends to give more considerable braking or acceleration commands. In contrast, the new reward function gives smaller and smoother braking or acceleration commands. The results show that the model obtained, considering the smoothness parameter, is more stable during unloading.

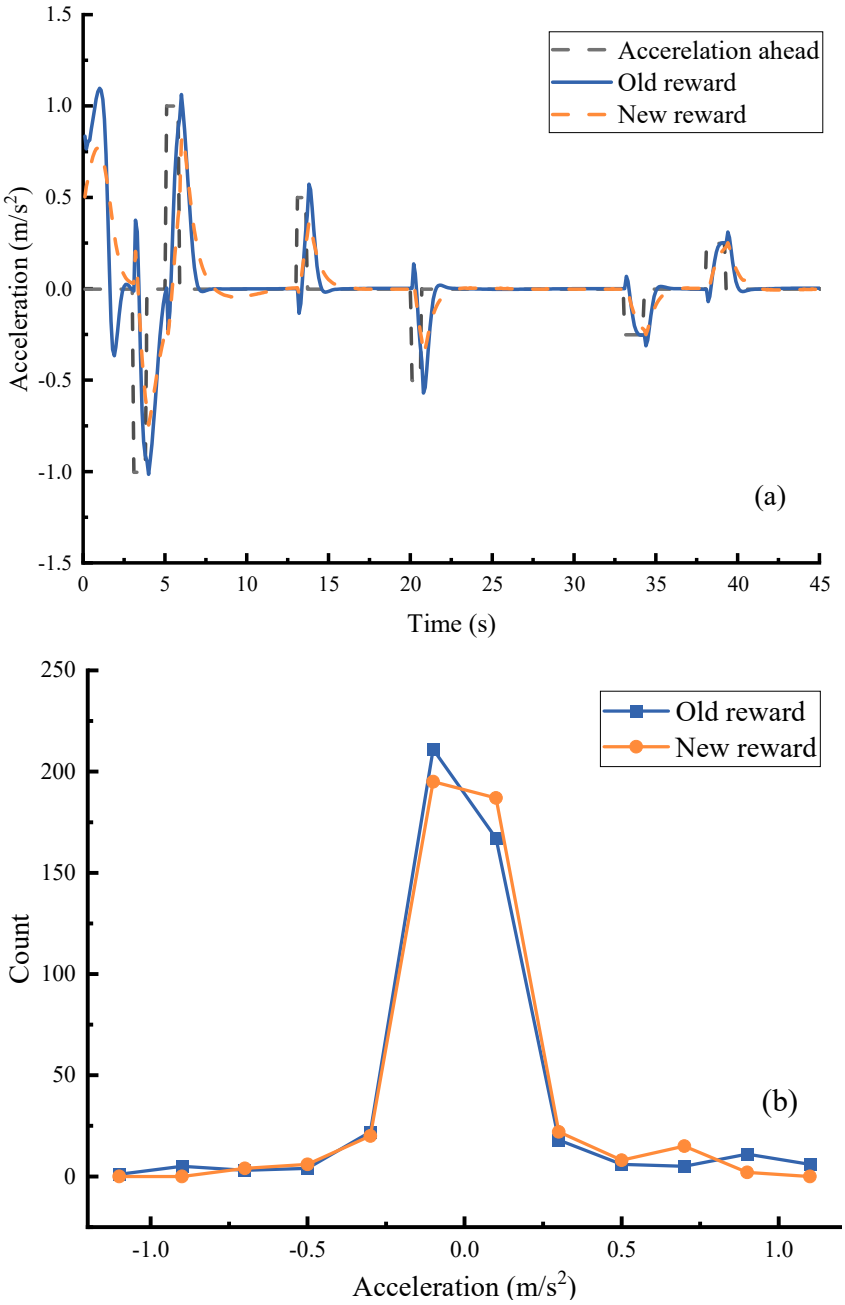


Figure 8. (**a**) Acceleration magnitude change graph; (**b**) Histogram of probability distribution of acceleration.

3.1.4. Comparison of Supervised Reinforcement Learning Algorithm and Linear (PID) Algorithm Simulation

In order to verify the performance of the supervised reinforcement learning algorithm proposed in the paper, a supervised reinforcement learning algorithm and linear algorithm simulation comparison test are carried out. Initially set at 1 m/s, the harvester began to accelerate at the 5th second until it reached 1.5 m/s, it then decelerated at the 15th second until it reached 1 m/s, and thereafter maintained a steady speed of 1 m/s, as shown in Figure 9. Figure 9a shows the acceleration comparison graph, in which the response times of the supervised reinforcement learning algorithm at 5 s and 15 s are 2.4 s and 2.6 s, respectively, and the response times of the linear algorithm are 2.8 s and 2.9 s, respectively. Compared with the linear algorithm, the supervised reinforcement learning algorithm responds faster in the initial phase of acceleration change, i.e., the supervised reinforcement learning algorithm has a more substantial response performance during the acceleration step. As shown in Figure 9b, the supervised reinforcement learning algorithm can keep up with the speed of the harvester faster when the speed of the harvester varies and has a maximum speed of 1.5 m/s compared to the linear algorithm, which has a maximum speed significantly more significant than the speed of the harvester. In summary, the supervised reinforcement learning algorithm responds faster to changes in acceleration and provides better velocity control than the linear algorithm.

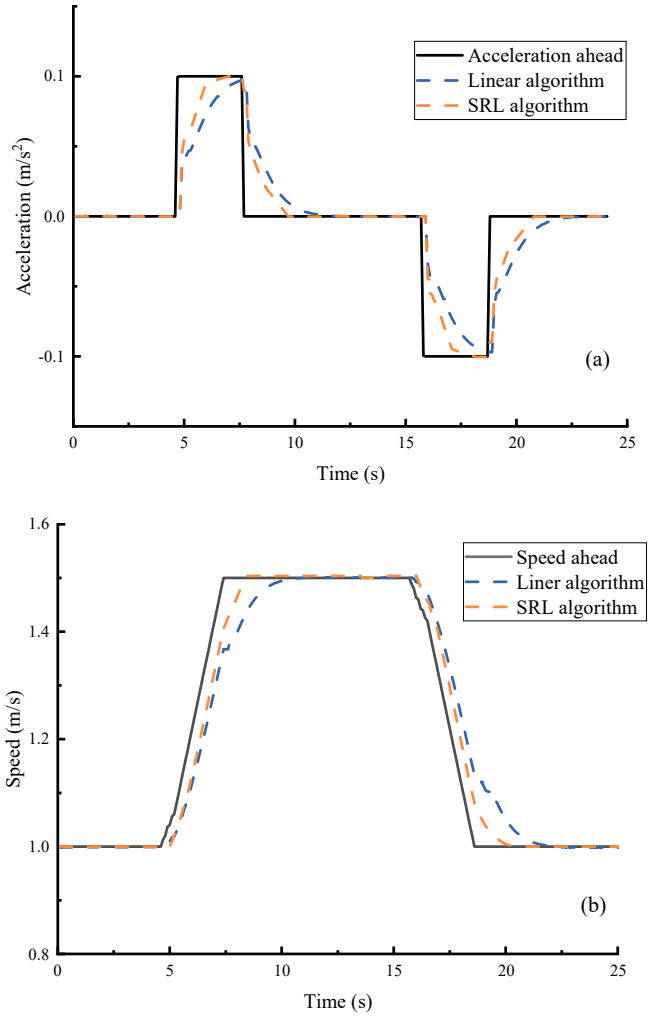


Figure 9. Step response plots of supervised reinforcement learning algorithm compared to linear algorithm. (**a**) Acceleration; (**b**) Speed.

3.2. Field Experiments

3.2.1. Preparation for Experimentation

Supervised reinforcement learning real-vehicle algorithms are developed using the MATLAB2019a/Simulink platform with visual programming. Calibration of throttle opening and brake pressure against the magnitude of acceleration was carried out by performing a variable speed test on the tractor to obtain a table of acceleration and brake pressure versus throttle opening conversion relationships. The throttle opening and brake pressure information are obtained by interpolating the current speed and desired acceleration look-up table. The throttle opening and brake pressure information are then converted into corresponding voltage values to be sent to the actuator to control the acceleration and deceleration of the tractor, which can be referred to in the paper.

The hardware devices used in the master–slave cooperative control framework designed in this paper include a positioning system, a model controller, and an inter-vehicle communication device. The positioning system is a combined GNSS/INS navigation system, which integrates gyroscopes and accelerometers, GNSS, magnetic field meters, barometers, and other devices. The data fusion filtering algorithm is embedded in the system's internal real-time operating system, which enables the system to output stable position data under different working conditions, with relative position accuracy less than 2.5 cm and attitude accuracy less than 1. The model controller is an intelligent controller developed based on an Infineon TC275 chip supporting a Simulink model design, which has peripheral interfaces of four CAN communication modules, two RS232 communication modules, eight low-side switches, two high-side switches, and one relay driver. The inter-vehicle communication device used is AMX-Lora vehicle-mounted wireless communication equipment. The following are the device's primary specifications: the communication interface supports two protocols, RS232 RS485; Baud rate default 9600, the maximum support of 115,200; this article uses 115,200 to achieve fast transmission while supporting parity bit, stop bit settings. Theoretical maximum communication distance is 1 km, and the maximum communication distance of Lora equipment can be up to 800 m under the condition of no obstruction through the test; the working frequency band is 410 MHZ~525 MHZ, which ensures the data transmission safely and stably; The data transmission mode supports stream mode and packet mode; the maximum transmission power is 20 dbm.

3.2.2. Experiment Process and Results

In Figure 10, the traveling following of the grain truck is shown when the harvester is traveling at a constant speed. The harvester moves for thirty seconds at a steady pace of 1.2 m/s. When the distance between the two vehicles reaches the distance set in the initial phase, the grain truck begins to accelerate. It follows up the speed of the harvester in a very short period, as shown in Figure 10a. As the speed of the harvester fluctuates back and forth up and down, the grain truck also responds slightly. However, the amount of change is small, proving that the algorithm can achieve a stable following of the speed of the two vehicles. Figure 10b demonstrates that the real distance of the two vehicles oscillates both above and below the intended distance, with a distance error contained within 0.26 m. The experimental results show that the algorithm can achieve the goal of grain trucks stably following the harvester to complete the following unloading operation under constant speed driving conditions, which verifies the validity of the master–slave cooperative control framework established in this paper.

In Figure 11, the traveling following of the grain truck is shown when the harvester travels at variable speed. In the first 26, the harvester travels at a constant speed of 0.4 m/s and accelerates in the 26th s until its speed becomes 1.2 m/s. Eventually, the speed was reduced to 0, as shown in Figure 11a. The desired following distance between the two vehicles was 3.1 m at a speed of 0.4 m/s, and the desired following distance was adjusted to 4.1 m when the vehicle speed was 1.2 m/s. Figure 11a shows that the lag time difference between the grain truck's speed response curve and the harvester's speed curve is concise. At the same time, the speed of the grain truck does not show any obvious over-response,

and it can follow the harvester's movement very well. In addition, from Figure 11b, the maximum following error at 0.4 m/s low speed is 0.21 m, and the maximum following error after the speed is stabilized at 1.2 m/s is 0.08 m, which has a good speed adaptability. The experimental results show that the algorithm can achieve the goal of the grain truck stably following the harvester to complete the following unloading operation under the variable speed driving condition, which verifies the adaptability of the master–slave cooperative control framework established in this paper.

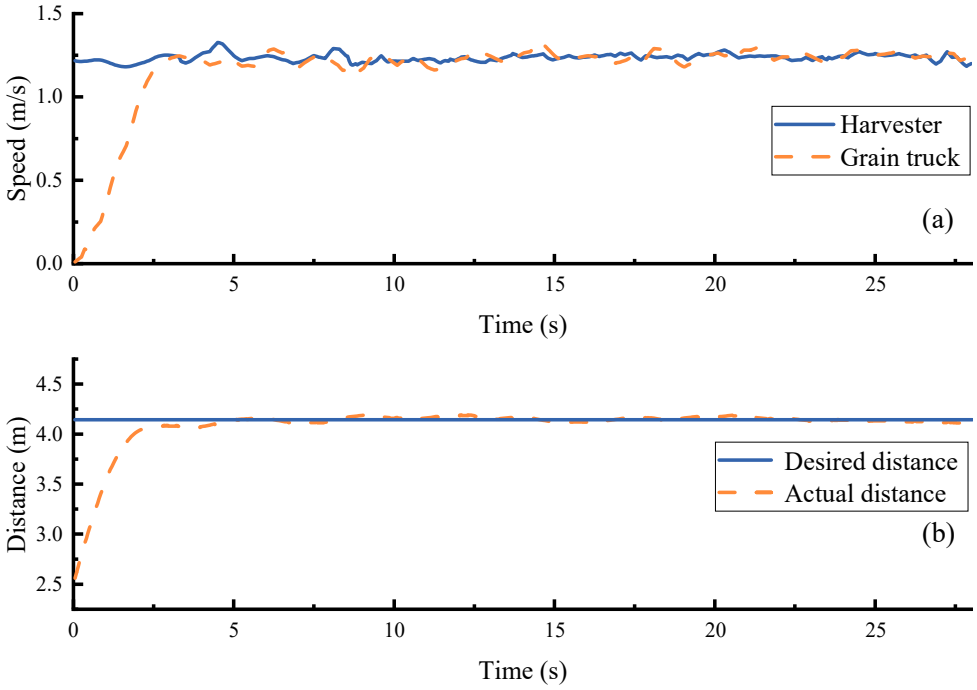


Figure 10. Follow figure at constant speed. (**a**) Speed; (**b**) Distance.

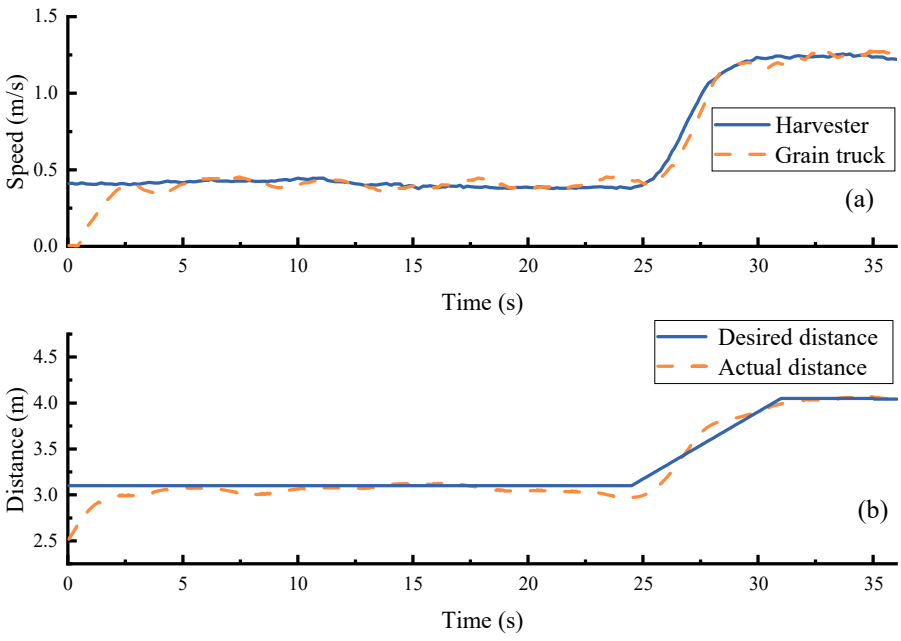


Figure 11. Variable speed driving following diagram. (**a**) Speed; (**b**) Distance.

4. Discussion

In Figure 10, when the harvester traveled at a constant speed of 1.2 m/s, it took only 6.5 s for the grain truck to reach the same speed as the harvester. When the speed of the two cars stabilized, after the harvester was affected by the ground at 4.6 s with speed fluctuations, the grain truck lagged behind by 1.6 s to respond quickly to the speed change of the harvester, and the speed fluctuation amplitude of the grain truck was significantly smaller than that of the harvester, with no overshooting. It shows that the algorithm has good corresponding speed and robustness.

From Figure 10, the harvester truck was traveling at 1.2 m/s, and the grain truck followed the harvester truck stably after 2.8 s of acceleration and speed adjustment; when the two trucks were following stably, the average absolute deviation of the speed of the grain truck was 0.028 m/s, the maximum deviation was 0.090 m/s, and the standard deviation was 0.032 m/s [41–43]. From Figure 11, the harvester truck accelerated after 5 s from 0.4 m/s to 1.2 m/s, although in this process, the grain truck, by the communication delay and response lag, lags 1 s after following the harvester truck to accelerate, but can follow the whole anthropomorphic harvester to accelerate. The harvester truck traveled at 0.4 m/s, the grain truck followed the harvester truck stably after 2.4 s of acceleration and speed adjustment, and the grain truck followed the harvester truck stably. When the two vehicles followed stably, the average absolute deviation of the grain truck's speed in the steady state was 0.023 m/s, the maximum deviation in the steady state was 0.081 m/s, and the standard deviation of the steady state was 0.026 m/s. When the harvester truck accelerated to 1.2 m/s and the two vehicles stabilized, the grain truck's average absolute deviation of the speed was 0.028 m/s, the maximum deviation was 0.063 m/s, and the standard deviation was 0.031 m/s. From the above, it can be seen that the algorithm designed in this paper can stably follow the speed change of the harvester truck, and the algorithm can complete the task of stable cooperative unloading of grain.

From Figure 10, the harvester truck was traveling at 1.2 m/s, and the grain truck could follow the harvester truck completely with the desired distance; when the two trucks are following stably, the maximum deviation of the actual distance between the two trucks in the steady state was 0.26 m, and the standard deviation of the steady state was 0.081 m. From Figure 11, the harvester truck was traveling at 0.4 m/s, when the two trucks were following stably, the maximum deviation of the actual distance between the two trucks in steady state was 0.20 m, and the standard deviation of the steady state was 0.042 m. When the harvester truck was traveling at 1.2 m/s and the two vehicles were following stably, the steady state maximum deviation of the actual distance between the two vehicles was 0.21 m, and the steady state standard deviation was 0.075 m. It can be seen in Figures 10 and 11 that when the speed of the harvester truck fluctuates, the distance change fluctuates less, and the actual distance and the desired distance are kept consistent. From the above, it can be seen that the algorithm designed in this paper can accurately control the actual distance between the two vehicles, to meet the requirements of collaborative accuracy [44], and the algorithm can complete the safe unloading task.

In summary, the supervised reinforcement learning-based wheat master–slave cooperative harvesting algorithm designed in this paper is able to adapt to different working conditions, and at the same time has good speed following effect and control accuracy. Future research continues to optimize the algorithm and improve the communication system, which will further enhance the performance of the master–slave cooperative control system and promote the development of an agricultural machinery cooperative operation.

5. Conclusions

1. In order to improve the algorithm training success rate, a supervisor trained on actual driving data was introduced into the AC reinforcement learning method to improve the performance and generalization of the model. Based on considering the smoothness of the unloading operation and the safety of unloading, a new reward function in the SRL algorithm is designed, and a supervised reinforcement learning-based multi-vehicle

master–slave cooperative unloading control algorithm is developed to realize the smooth and safe unloading of the grain truck.

2. Simulation comparison experiments show that the reinforcement learning algorithm with the introduction of a supervisor can converge stably. The supervised reinforcement learning algorithm that combines operational smoothness and unloading safety can ensure that the grain truck unloads grain smoothly and safely. The proposed supervised reinforcement learning algorithm responds to acceleration changes faster and has a more accurate speed adaptability than the linear algorithm. The field operation experiments show that when the harvester travels at a constant speed of 1.2 m/s, the following error is no more than 0.18 m, and the steady state standard deviation is 0.081 m. When the harvester is traveling at a speed of 0.4 m/s at a constant speed and then the speed increases to 1.2 m/s, the following errors at these two speeds are no more than 0.1 m and 0.08 m, respectively, and the steady state standard deviation is 0.042 m and 0.075 m. The algorithm can achieve the master–slave unloading control for grain trucks and harvester workshops, and it can be adapted to different operating conditions.

3. Harvesting operations are mainly carried out at a constant speed with slow traveling speeds. The harvesting operation scenarios have become more complex with the increasing harvester intelligence and harvesting requirements. Therefore, improving master–slave cooperative unloading control is a future research direction. The algorithm research ideas and results presented in this paper provide a reference method for master–slave cooperative control and optimization in the field.

Author Contributions: Conceptualization, Z.M. and C.Z. (Chao Zhang); methodology, Z.M.; software, Z.M.; validation, Z.M., C.Z. (Chao Zhang), and H.W.; formal analysis, Z.M.; investigation, H.W.; resources, Z.M.; data curation, C.Z. (Chao Zhang); writing—original draft preparation, Z.M.; writing—review and editing, Z.M. and C.Z. (Chao Zhang); visualization, C.Z. (Chao Zhang); supervision, H.Y.; project administration, C.Z. (Chunjiang Zhao); funding acquisition, W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant number 2023YFD2301503), Infrared-enhanced sensor-based growth monitoring model and application demonstration for facility horticulture crops (grant number NK202302020205), and the Central Guidance on Local Science and Technology Development Fund of Hebei Province (grant number 236Z7203G).

Data Availability Statement: The data presented in this study are available on request from the first author.

Acknowledgments: The authors gratefully acknowledge the editors and anonymous reviewers for their constructive comments on our manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lan, Y.B.; Zhao, D.N.; Zhang, Y.F.; Zhu, J.K. Exploration and development prospect of eco-unmanned farm modes. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 312–327. [CrossRef]
2. Luo, X.W.; Liao, J.; Hu, L.; Zhou, Z.Y.; Zhang, Z.G.; Zang, Y.; Wang, P.; He, J. Research progress of intelligent agricultural machinery and practice of unmanned farm in China. *J. South China Agric. Univ.* **2021**, *42*, 8–17+15. [CrossRef]
3. Liu, C.L.; Lin, H.Z.; Li, Y.M.; Gong, L.; Miao, Z.H. Analysis on status and development trend of intelligent control technology for agricultural equipment. *Trans. Chin. Soc. Agric. Eng.* **2020**, *51*, 1–18.
4. Charania, I.; Li, X.R. Smart farming: Agriculture's shift from a labor intensive to technology native industry. *Internet Things* **2020**, *9*, 100142. [CrossRef]
5. Zhao, C.J.; Li, J.; Feng, X.; Ma, C. Reflections on the Development of Intelligent Agricultural Machinery and Equipment. *Issues Agric. Econ.* **2023**, *44*, 4–12. [CrossRef]
6. Shen, Y.; Zhang, Y.F.; Liu, H.; He, S.W.; Feng, R.; Wan, Y.L. Research Review of Agricultural Equipment Automatic Control Technology. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 1–18. [CrossRef]
7. Li, D.L.; Li, Z. System analysis and development prospect of unmanned farming. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 1–12. [CrossRef]

8. Luo, X.; Hu, L.; He, J.; Zhang, Z.; Zhou, Z.; Zhang, W.; Liao, J.; Huang, P. Key technologies and practice of unmanned farm in China. *Trans. Chin. Soc. Agric. Eng.* **2024**, *40*, 1–16. [CrossRef]
9. Liu, Z.P.; Dhamankar, S.; Evans, J.T.; Allen, C.M.; Jiang, C.; Shaver, G.M.; Etienne, A.; Vyn, T.J.; Puryk, C.M.; McDonald, B.M. Development and experimental validation of a system for agricultural grain unloading-on-the-go. *Comput. Electron. Agric.* **2022**, *198*, 107005. [CrossRef]
10. Liu, Z.P.; Dhamankar, S.; Evans, J.T.; Allen, C.M.; Jiang, C.; Shaver, G.M.; Etienne, A.; Vyn, T.J.; Puryk, C.M.; McDonald, B.M. Automation of Agricultural Grain Unloading-on-the-go. *IFAC-Pap.* **2022**, *55*, 248–253. [CrossRef]
11. Iida, M.; Kudou, M.; Ono, K.; Umeda, M. Automatic following control for agricultural vehicle. In Proceedings of the 6th International Workshop on Advanced Motion Control. Proceedings (Cat. No.00TH8494), Nagoya, Japan, 30 March–1 April 2000; pp. 158–162.
12. Cao, R.Y.; Zhang, Z.Q.; Li, S.C.; Zhang, M.; Li, H.; Li, M. Multi-machine cooperation global path planning based on a-star algorithm and bezier curve. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 548–554. [CrossRef]
13. Li, S.C.; Cao, R.Y.; Wei, S.; Ji, Y.H.; Zhang, M.; Li, H. Development of multi-vehicle cooperative navigation communication system based on TD-LTE. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 45–51+65.
14. Kan, X.Y.; Thayer, T.C.; Carpin, S.; Karydis, K. Task planning on stochastic aisle graphs for precision agriculture. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3287–3294. [CrossRef]
15. Cao, R.Y.; Li, S.C.; Wei, S.; Ji, Y.H.; Zhang, M.; Li, H. Remote monitoring platform for multi-machine cooperation based on Web-GIS. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 52–57+14. [CrossRef]
16. Zhai, Z.Q.; Wang, X.Q.; Wang, L.; Zhu, Z.; Du, Y.; Mao, E. Collaborative path planning for autonomous agricultural machinery of master-slave cooperation. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 542–547. [CrossRef]
17. Cao, R.Y.; Li, S.C.; Ji, Y.H.; Xu, H.Z.; Zhang, M.; Li, M.Z. Multi-machine cooperation task planning based on ant colony algorithm. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 34–39. [CrossRef]
18. Lee, K.; Ehsani, R.; Schueller, J. Forward movement synchronization of two vehicles in parallel using a laser scanner. *Appl. Eng. Agric.* **2007**, *23*, 827–834. [CrossRef]
19. Zhang, C.; Noguchi, N.; Yang, L.L. Leader–follower system using two robot tractors to improve work efficiency. *Comput. Electron. Agric.* **2016**, *121*, 269–281. [CrossRef]
20. Luo, C.M.; Mohsenimanesh, A.; Laguë, C. Synchronous Tracking Control for Agricultural Wide-Span Implement Carrier (WSIC). *Trans. ASABE* **2018**, *61*, 873–883. [CrossRef]
21. Li, S.C.; Xu, H.Z.; Ji, Y.H.; Cao, R.Y.; Zhang, M.; Li, H. Development of a following agricultural machinery automatic navigation system. *Comput. Electron. Agric.* **2019**, *158*, 335–344. [CrossRef]
22. Bai, X.P.; Hu, J.T.; Wang, Z. Slave positioning method for cooperative navigation of combine harvester group based on visual servo. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 59–68. [CrossRef]
23. Bai, X.P.; Wang, Z.; Hu, J.; Gao, L.; Xiong, F. Harvester group corporative navigation method based on leader-follower structure. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 14–21. [CrossRef]
24. Xu, G.F.; Chen, M.Z.; Miao, H.Q.; Yao, W.Y.; Diao, P.S.; Wang, W.J. Following operation control method of farmer machinery based on model predictive control. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 11–20. [CrossRef]
25. Chen, J.; Zhu, Z.X.; Liang, N.C.; Takeda, J. Automatic On-tracking Control of Farm Vehicle Based on Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2007**, 131–133+121. [CrossRef]
26. Wang, Y.W.; Guo, Z.J.; Wu, J.B.; Fu, S.Z. Research on vehicle adaptive cruise control based on BP neural network working condition recognition. *JoE* **2022**, *2022*, 132–147. [CrossRef]
27. Gautron, R.; Maillard, O.-A.; Preux, P.; Corbeels, M.; Sabbadin, R. Reinforcement learning for crop management support: Review, prospects and challenges. *Comput. Electron. Agric.* **2022**, *200*, 107182. [CrossRef]
28. Xiong, J.T.; Li, Z.H.; Chen, S.M.; Zheng, Z.H. Obstacle avoidance planning of virtual robot picking path based on deep reinforcement learning. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 1–10. [CrossRef]
29. Shao, J.K.; Zhao, X.; Yang, J.; Zhang, W.; Kang, Y.; Zhao, X. Reinforcement learning algorithm for path following control of articulated vehicle. *Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 376–382. [CrossRef]
30. Zhang, W.; Zhang, Z.; Luo, X.; He, J.; Hu, L.; Yue, B. Position-velocity coupling control method and experiments for longitudinal relative position of harvester and grain truck. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 1–11. [CrossRef]
31. Wei, S.Y.; Zou, Y.; Li, X.L.; Wang, W.W. A research on connected cruise control for vehicles. *Automot. Eng.* **2018**, *40*, 290–296. [CrossRef]
32. Wang, Y.P.; Zheng, K.X.; Tian, D.X.; Duan, X.T.; Zhou, J.S. Pre-training with asynchronous supervised learning for reinforcement learning based autonomous driving. *Front. Inf. Technol. Electron. Eng.* **2021**, *22*, 673–686. [CrossRef]
33. Zhu, J.Z.; Wang, C.; Li, X.K.; Dong, Y.C.; Zhang, H.L. A deep reinforcement learning based discrete state transition algorithm for fuzzy flexible job shop scheduling. *J. Beijing Univ. Aeronaut. Astronaut.* **2023**, 1–15. [CrossRef]
34. Zhao, D.B.; Hu, Z.H.; Xia, Z.P.; Alippi, C.; Zhu, Y.H.; Wang, D. Full-range adaptive cruise control based on supervised adaptive dynamic programming. *Neurocomputing* **2014**, *125*, 57–67. [CrossRef]
35. Wang, Y.D.; Dong, L.; Sun, C.Y. Cooperative control for multi-player pursuit-evasion games with reinforcement learning. *Neurocomputing* **2020**, *412*, 101–114. [CrossRef]

36. Wei, M.X.; Yang, J.W.; Chen, K.; Wang, Z.H.; Sha, C. Vehicle longitudinal following based on improved brain emotional learning model. *J. Jilin Univ. (Eng. Technol. Ed.)* **2022**, *52*, 2994–3005. [CrossRef]
37. Wei, S.Y.; Zou, Y.; Zhang, T.; Zhang, X.D.; Wang, W.W. Design and experimental validation of a cooperative adaptive cruise control system based on supervised reinforcement learning. *Appl. Sci.* **2018**, *8*, 1014. [CrossRef]
38. Cui, Z.; Hu, J.; Yu, Y.; Cao, G.; Zhang, H.; Chai, X.; Chen, H.; Xu, L. Automatic grain unloading method for track-driven rice combine harvesters based on stereo vision. *Comput. Electron. Agric.* **2024**, *220*, 108917. [CrossRef]
39. Tan, H.C.; Zhang, H.L.; Peng, J.K.; Jiang, Z.X.; Wu, Y.K. Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space. *Energy Convers. Manag.* **2019**, *195*, 548–560. [CrossRef]
40. Zhu, M.X.; Wang, Y.; Pu, Z.; Hu, J.; Wang, X.S.; Ke, R. Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102662. [CrossRef]
41. Ding, F.; Zhang, W.; Luo, X.; Hu, L.; Zhang, Z.; Wang, M.; Li, H.; Peng, M.; Wu, X.; Hu, L.; et al. Gain self-adjusting single neuron PID control method and experiments for longitudinal relative position of harvester and transport vehicle. *Comput. Electron. Agric.* **2023**, *213*, 108215. [CrossRef]
42. Ding, F.; Zhang, W.; Luo, X.; Zhang, Z.; Wang, M.; Li, H.; Peng, M.; Hu, L. Design and Experiment for Inter-Vehicle Communication Based on Dead-Reckoning and Delay Compensation in a Cooperative Harvester and Transport System. *Agriculture* **2022**, *12*, 2052. [CrossRef]
43. Wang, H.; Wang, G.M.; Luo, X.W.; Zhang, Z.G.; Gao, Y.; He, J.; Yue, B.B. Path tracking control method of agricultural machine navigation based on aiming pursuit model. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 11–19. [CrossRef]
44. Ahmed, S.; Qiu, B.J.; Kong, C.W.; Xin, H.; Ahmad, F.; Lin, J.L. A Data-Driven Dynamic Obstacle Avoidance Method for Liquid-Carrying Plant Protection UAVs. *Agronomy* **2022**, *12*, 873. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.