

Review Article

Kai Feng, Lingzhi Yang*, Dongfeng He, Shijing Lin, and Buxin Su

A study on deep reinforcement learning-based crane scheduling model for uncertainty tasks

<https://doi.org/10.1515/htmp-2022-0040>

received January 19, 2022; accepted May 16, 2022

Abstract: Aiming at the crane scheduling problem for uncertainty tasks in multi-crane scheduling situation, this article proposes a deep reinforcement learning-based crane scheduling modeling method that is not dependent on mathematical planning and has certain generality. First, the crane scheduling process is integrated into deep reinforcement learning framework in which the orbit space of the crane and the transportation task is environmental information and crane is the intelligent agent. Second, the interactive mode between reinforcement learning algorithm and environment is adjusted to adapt to the combined learning of multi-crane scheduling model. Last, the crane scheduling model for uncertainty tasks is constructed by optimizing reward discount factor, learning rate, and reward function intensive mode. Testing of the model is carried out based on practical crane scheduling in one steelmaking workshop. Scheduling proposal is generated and all crane tasks are completed within the planned time, which verifies the feasibility of this model. Results show that compared with manual scheduling plan, the scheduling proposal based on the new model reduces total task completion time by 11.52%, time of collision of crane routes decreases by 57.14%, and negative crane transportation distance shortens by 55.26%. The high efficiency of the scheduling model is therefore verified.

Keywords: crane scheduling, deep learning, reinforcement learning, steelmaking workshop, reward function

1 Introduction

Crane, like the bridge crane operated in overhead track, is widely applied in situations such as port terminals, large warehouses, and industrial workshops. The research works on cranes are generally in two aspects. One focuses on security, monitoring, and analyzing the useful life, status, and abnormality of the crane equipment. Another aspect is about its efficiency, optimizing the crane transportation, and scheduling process. In recent years, with the increasing level of automation and informatization in logistics and warehousing in various industrial situations, more and more researchers and engineers pay attention to the efficiency improvement in crane scheduling in respective scenarios.

According to the number of cranes on the tracks, there are two scenarios: single crane scheduling and multi-crane scheduling. Single crane scheduling refers to only one crane operating in the same crane track. The scheduling efficiency in this situation depends on the execution sequence of transportation tasks. Multi-crane scheduling indicates two or more cranes operating simultaneously in the same crane track. The scheduling efficiency is mainly affected by two factors: matching of cranes and tasks, and crane movement route planning. Currently, most research works in efficiency focus on the second scenario.

In the situation of multi-crane scheduling, categorized by the features of transportation tasks, there are scheduling for certainty tasks and uncertainty tasks. The former means the transportation tasks within certain period of time can be accurately predicted, and thus task plans can be drawn up and respective crane scheduling plans can be generated to instruct the completion of transportation tasks. While in the latter scenario, the tasks cannot be accurately predicted, therefore only reactive crane scheduling method can be generated based on

* **Corresponding author: Lingzhi Yang**, School of Minerals Processing and Bioengineering, Central South University, Changsha 410083, China, e-mail: yanglingzhi@163.com

Kai Feng, Dongfeng He: School of Metallurgy and Ecology Engineering, University of Science and Technology Beijing, Beijing 100083, China

Shijing Lin: Midea Group Co., Ltd, Foshan, 528311, China

Buxin Su: China Metallurgical Industry Planning and Research Institute, Beijing 100013, China

instantly emerged tasks. By far, there are wide research works on crane scheduling for certainty tasks, but relatively only fewer research works are available on crane scheduling for uncertainty tasks.

This article targets at the crane scheduling for uncertainty tasks in multi-crane scheduling situation. A deep reinforcement learning-based crane scheduling model that is not dependent on mathematical planning and has certain generality is proposed. First, the crane scheduling model is constructed under the deep reinforcement learning framework, the orbit space of the crane and the transportation task is the environmental information and crane is the intelligent agent. Second, the network structure, reward function, and decision rule of the scheduling model is designed and adjusted, the model is trained and optimized. Last, testing is carried out based on actual crane scheduling in one steelmaking workshop. Scheduling plan is generated using the model, and comparison with actual crane scheduling plan is made. The feasibility and efficiency of the method is proved.

2 Literature review

Crane, as one type of conveyance for heavy goods in fixed area, is widely applied in various industries. Typical situations include transit of shipment containers in port terminals [1], transportation of yard containers [2], production and warehouse workshops in manufacturing enterprises [3], and so on.

In some of the application situations, one crane can complete all tasks since there are relatively less transportation tasks. Under this situation, the focus of crane scheduling research is on optimizing the execution sequence to complete the tasks more efficiently. For instance, only one crane is needed to transport high temperature slab in the production workshop of hot rolled slab in iron and steel enterprises. Therefore, a scheduling model was constructed based on hybrid metaheuristic algorithm with the target to achieve lowest total temperature drop of transported slab [4]. Similarly, in the warehouses of steel coil products, some enterprises only adopt one crane to complete the stacking and transportation of coil products. A scheduling model was constructed with the target to achieve minimum crane scheduling time [5].

In other application situations, large number of transportation tasks are to be completed in a big space. In such situations, constructing different crane scheduling

methods and plans according to different constraints of business situation and transportation task features becomes the major content of crane scheduling research. For example, for crane scheduling of yard containers, a scheduling plan with the target of minimum waiting time was generated using heuristic method and its effectiveness is verified with actual data [6]. For crane scheduling of port containers, an effective method was proposed for quay crane scheduling of multiple hatches vessel considering double-cycling strategy to improve the operation efficiency and reduce the risk of delay [7]. Focusing on the high randomness of loading and unloading operations in port containers, separate modeling for crane allocation and detailed scheduling plan were made and computed [8]. For crane scheduling of warehouse workshops, due to the constraints of space and security, an integrated decision model was created to combine the location allocation of goods, crane distribution, and route planning [9]. For crane scheduling of hot rolled slab in iron and steel enterprises, scheduling model was constructed with consideration of actual constraints of time, space, and hot rolling plan. Simulation testing was carried out using actual production data and Memetic algorithm [10].

According to the methods of constructing crane scheduling model, research works can be categorized to modeling based on mathematical programming and modeling based on intelligent algorithm.

In modeling based on mathematical programming, mixed integer programming model is widely studied. For instance, for the coordination issue between guided vehicle and cranes in the scheduling of containers, a mixed integer programming model was created with the target of minimized completion time for maximum task batches. Its efficiency was compared with heuristic method, genetic algorithm and particle swarm algorithm [11]. For the crane scheduling of port containers and yard containers, a mixed integer programming model was constructed targeting at minimum task completion time and truck travelling time. A hybrid particle swarm optimization algorithm was proposed integrating neighborhood search of tabu list and heuristic preprocessing approach [12]. Aiming at the stacking problem in crane scheduling for the storage workshop of finished products, a mixed integer programming model was established to avoid such situations to the largest extent [13]. For the crane scheduling problem in production workshops in iron and steel enterprises, targeting at minimum total tasks completion time, an integer programming model was proposed by setting up task caching method and sequential

execution strategy, and computing using the separation heuristic algorithm [14].

In terms of modeling based on intelligent algorithm, typical methods include heuristic algorithm, simulation method and genetic algorithm, etc. For example, for crane scheduling in steelmaking workshops, based on production plan and technical requirements of actual production, optimized crane scheduling plan was obtained by integrating heuristic rules and simulation [15]. In respect of multi-machine multi-task crane scheduling, with the consideration of space constraints and priority of transportation tasks, a simulation model was proposed integrating immune algorithm and genetic algorithm and its industry application was assessed [16]. A crane scheduling modeling method mixing time Petri net was raised, shortening the full production cycle of steelmaking and continuous casting to the maximum [17]. In the situation of scheduling port containers with uncertain loading and unloading time, a scheduling model was created and a feasible scheduling plan was obtained by adopting mixed integer programming model and genetic algorithm search [18].

Besides above studies, to reduce the difficulty of applying crane scheduling model, research works were also conducted on how to simplify the modeling process and improve computing efficiency under the precondition of achieving same scheduling effect. Examples include: in the situation of port container scheduling, considering the constraints of berth length, a scheduling model was created by integrating accurate solution and approximate solution, and the computation efficiency can be effectively improved [19]. In terms of multi-crane scheduling under non-cross constraints, a method combining the backtrack search and pruning strategy was constructed and crane distribution plan can be generated quickly [20]. To reduce the complexity of scheduling model, in the situation of yard containers, a step-by-step decision-making approach was adopted and mathematical modeling was established. Multi-level genetic algorithm is applied in computation [21]. In the scenario of port container scheduling, a two-phase random planning scheduling model was proposed to tackle the feature of uncertain release time of transportation tasks. Genetic algorithm and heuristic rule algorithm were adopted in computation [22]. In terms of interference in multi-crane scheduling in steel manufacturing, a heuristic algorithm that can achieve suboptimal solution within short time was presented to create crane scheduling model, with its feasibility verified in computer experiment [23]. To simplify the scheduling objects in port containers, container movement was used to mark the crane tasks and symbolize crane distribution

and movement process. Respective mathematical modeling was created [24]. Dohn and others divided the slab yard scheduling into two parts: planning and scheduling, separate modeling and computation were carried out [25].

By far, the research works on multi-crane scheduling mainly focuses on constructing the model using mathematical programming or intelligent algorithm. Such models can achieve good scheduling effects in certain application situations. However, the models or methods lack generality and can hardly be applied to other situations. They also cannot properly deal with the scheduling problem for uncertainty tasks. To develop a scheduling model targeting uncertainty tasks and having certain generality, this article proposes a solution based on deep reinforcement learning. According to this method, neural networks are trained using historic data in the specific situation and the crane scheduling model for this situation is obtained. Feasible and highly efficient crane scheduling plan can be generated for uncertainty tasks.

Deep reinforcement learning is an artificial intelligence method to solve the sequential optimization decision-making problem. Crane scheduling just falls into such type. As early as 2,000, Arai [26] from Carnegie Mellon University in the United States proposed the use of multi-agent reinforcement learning method to solve the collision problem caused by the uncertainty of crane scheduling task. In the meanwhile, for problems with same characteristics in other fields, reinforcement learning has been widely studied. For example, aiming at the uncertainty of external environment and non-linearity of control input for underwater vehicle, an adaptive control model is constructed using reinforcement learning [27]. For the problem of tracking moving targets in dynamic environment, a navigation model is created by reinforcement learning and genetic network programming based on the predicted collision characteristics [28]. Regarding the traffic signal adaptive control problem in city management, multiple reinforcement learning frameworks are used for parallel training and building an integrated control model [29]. In recent years, with the rapid development of deep learning theory, deep reinforcement learning that combines reinforcement learning has obtained a stronger ability to solve problems [30–32]. It has rapidly become an important research method in many fields. The most typical application of deep reinforcement learning framework is AlphaGo Zero [33], a model targeting the multi-step continuous decision optimization in the process of Go game. The solution is far more optimized than human experts and AlphaGo Lee and AlphaGo master based on supervised deep learning.

3 Deep reinforcement learning-based crane scheduling method

Crane scheduling for uncertainty tasks in multi-crane scheduling situation generally has the following common features:

- (1) Several cranes distributed across the same cross-region and equipped in same track lead to certain sequence and cranes cannot overstep each other to carry out the transportation tasks.
- (2) The tonnages and functions of multiple cranes are basically consistent, cranes can execute other transportation tasks.
- (3) The generation of uncertainty tasks is not fully random due to the constraints of actual business logic. Rather, there is a certain probability correlation among the time domain, space domain, or the tasks, therefore it has certain Markov property.
- (4) The starting and ending point of crane transportation tasks remains within the cross-region, the objective of crane scheduling model is to complete the transportation tasks more efficiently.

With the above analysis, it can be concluded that the crane scheduling of uncertainty tasks in multi-crane situation belongs to sequential decision-making problem within a limited space. Deep learning can realize the extraction of effective information within the space and reinforcement learning searches for the optimal strategy for sequential decision-making problem. The combination of deep learning and reinforcement learning can provide an effective solution for the crane scheduling for uncertainty tasks in multi-crane situation.

3.1 Basic principles of deep reinforcement learning

Deep reinforcement learning is a method that combines the perceptual ability from deep learning and decision-making from reinforcement learning. It can realize a systematic and integrated intelligent algorithm framework from perception to decision-making.

Deep learning adopts convolutional neural network, it is a type of feedforward neural network which contains convolutional calculation and deep structure, normally composed of convolutional layer, pooling layer, and full link layer. In reinforcement learning, intelligent agents with learning ability have interactive trial and error with their environment, allowing the continuous learning and

achievement of the objective. General process of reinforcement learning is as below:

- (1) The intelligent agent observes the status of environment s_t at moment t and selects action a_t based on current strategy π ;
- (2) Affected by action a_t , the environment status changes to s_{t+1} , assessment on the selected action of the intelligent agent is carried out and reward r_t is given based on one type of reward function R ;
- (3) The intelligent agent adjusts the strategy according to the reward feedback, and the process returns to step (1).

The intelligent agent finally obtains an optimal strategy π^* by constantly repeating these steps, and reaches the highest accumulated reward value by end of the cycle.

The Q -learning algorithm adopted in deep learning is the classic algorithm in computing optimal strategy π^* . The core is the status-action value function $Q^\pi(s, a)$ through learning and iterating. The function describes the value of each action a under status s . When achieving the optimal function $Q^\pi(s, a)$, the optimal strategy refers to the action a_t that leads to maximum value of $Q^{\pi^*}(s_t, a_t)$ under status s_t . The iteration and learning of $Q^{\pi^*}(s_t, a_t)$ is shown as below:

$$Q_{k+1}^\pi(s_t, a_t) = Q_k^\pi(s_t, a_t) + \alpha[r_t + \gamma \max_{a_{t+1}} Q_k^\pi(s_{t+1}, a_{t+1}) - Q_k^\pi(s_t, a_t)], \quad (1)$$

where k is the times of iteration, s_t is the status at moment t , a_t is the action at moment t , r_t is the immediate reward executing action a_t under status s_t . Similarly, s_{t+1} and a_{t+1} refer to the status and action at the next moment. $\alpha \in (0, 1)$ is the learning rate and $\gamma \in (0, 1)$ is the reward discount factor.

3.2 Deep reinforcement learning-based crane scheduling model

The deep reinforcement learning-based crane scheduling model views each crane as single intelligent agent and the respective workshop as the environment. The action of the intelligent agent is the action of crane, the status it observes is the status of task information and status of all cranes within the cross-region. Instructed by reward function that is designed through objective function, cranes can complete the task highly efficiently. The formation process of reinforcement learning mechanism is shown as Figure 1.

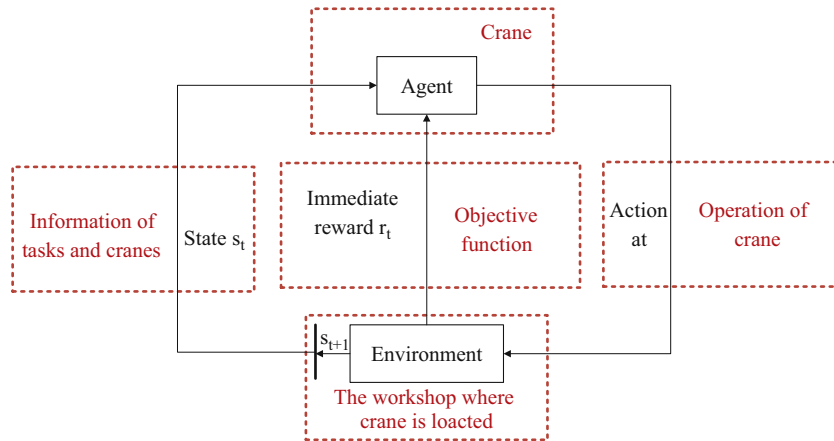


Figure 1: Reinforcement learning mechanism of crane scheduling model.

- (1) Environment is the scene that the intelligent agent interacts within, it is normally an abstraction of the actual environment, including the status space, action space, and reward function.

Status space is the set of potential status the environment can occur, which needs to contain all feature information necessary for decision-making. To reduce the complexity of status space, for crane scheduling problem, the cross-region space is abstracted to a two-dimensional matrix. The status space is symbolized through the status information marked by different values in the matrix. This study selected a 3×30 matrix to symbolize status space. The first row indicates the beginning position of transportation task, the second row refers to the numbering and status information of cranes, and the third row shows the ending position of the transportation task. Sample of the status space matrix is shown as Figure 2.

Such design of status space neglects the crane movement along the main beam direction, since such movement takes much less time and distance compared to the orbital direction and therefore, barely has any impact on crane scheduling.

Action space is the set of selectable actions generated by the interaction between intelligent agents and environment. In crane scheduling, cranes normally have seven actions, which are moving left/right along orbital direction, moving forward/backward along main beam direction,

standstill, lifting, and putting down. Since the selected status space in this study has neglected the movement along main beam direction, the action space is simplified to 5 instead of 7, represented by 0, 1, 2, 3, and 4, respectively.

Reward function is the assessment rule for the actions of intelligent agent. It is an important instruction signal for the intelligent agent to learn and improve and should be designed according to the objective function of the problems. To obtain a practical and efficient crane scheduling model, minimum task completion time is set as the objective and the reward function is designed as formula (2).

$$R = \begin{cases} 0, & \text{When the action doesn't end,} \\ f(t), & \text{At the end of the action.} \end{cases} \quad (2)$$

where $f(t)$ is the function negatively correlated to the task completion time t , which means, when t is smaller, and $f(t)$ is higher.

As the reward function is designed to have reward only after completing the task, there can be a problem of reward sparsity. Therefore, intensive improvement for reward function is required, that is to supplement some reward value to the five actions left/right movement along the orbital direction, standstill, lifting, and putting down. In the meantime, add negative reward for abnormal situations like collision and moving out of the reasonable range.

- (2) Intelligent agent consists of neural network structure and reinforcement learning algorithm.

	0	1	2	3	4	5	6	7	26	27	28	29
Starting stations	0	0	1	0	0	0	0	0	0	0	0	0
Cranes	0	0	0	1	0	0	2	0	0	0	0	0
Ending stations	0	0	0	0	0	1	0	0	0	0	0	0

Figure 2: 3×30 matrix of status space in crane scheduling.

Neural network structure adopts convolutional neural network in deep learning. For crane scheduling issue, the 3×30 matrix of status space in crane scheduling is the input for convolutional neural network. The five actions in action space are its output. Detailed neural structure is shown in Figure 3.

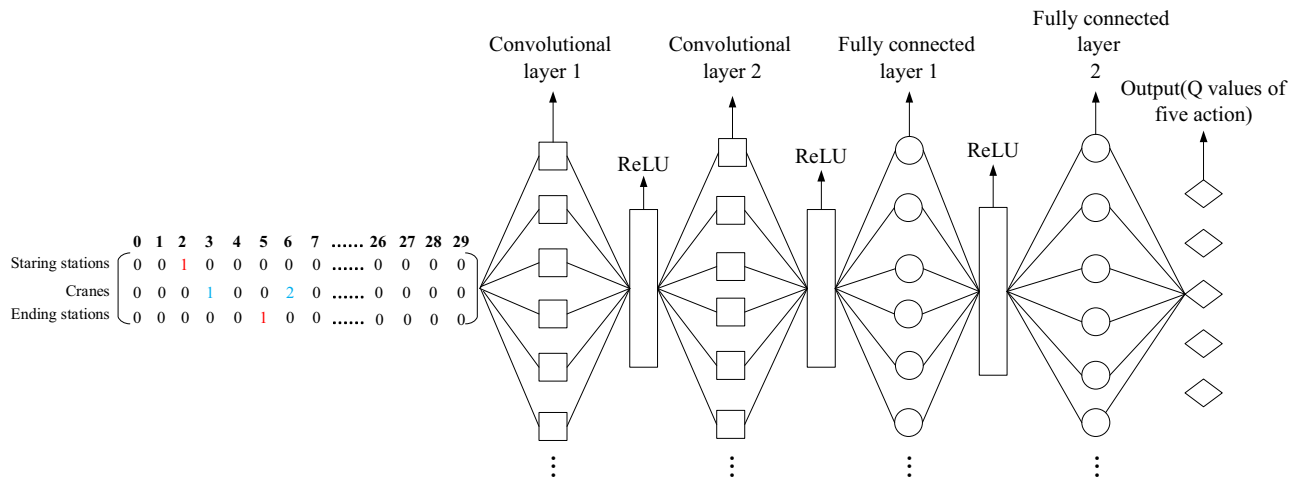


Figure 3: Sketch of neural network structure of intelligent agent.

The convolutional neural network includes two convolution layers and two full link layers. The convolution kernel has a size of 3×3 , the number of neurons in full link layer is 1,024, and activation function uses the ReLU function.

In crane scheduling, specific convolutional neural network needs to be designed for each crane. How to coordinate multiple neural networks to complete the training simultaneously is the key in the research of reinforcement learning algorithm in crane scheduling. This study made improvement in redesigning the interaction between multiple Q network and the environment on the basis of Deep Q-Learning (DQN) algorithm. Taking two cranes as example, the interaction is shown as Figure 4.

In Figure 4, 1–9 is the Q network training step sequence for two cranes; s_t^1 is the environment status of

crane 1 observed in moment t ; s_t^2 is the environment status of crane 2 observed in moment t ; s_{t+1}^1 is the environment status of crane 1 observed in moment $t + 1$; a_t^1 is the action of crane 1 in moment t ; a_t^2 is the action of crane 2 in moment t ; r_t^1 is the immediate reward to crane 1 in moment t ; r_t^2 is the immediate reward to crane 2 in moment t ; r_t is the final immediate reward to two cranes in moment t .

The training process of multi-crane is similar to that of two cranes. Within one unit of time, multiple cranes select one action according to the environment and act upon it sequentially. The sum of respective reward value is the immediate reward of multiple network models. When there are conflicts in the routes, they will choose the optimal avoiding method to get the highest reward value, since the immediate reward is the sum of rewards these cranes can get individually.

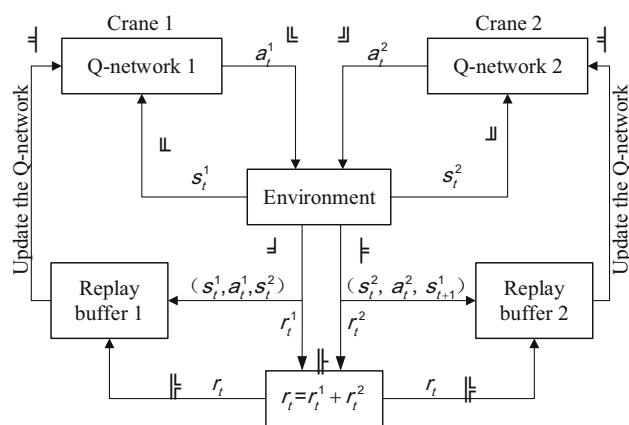


Figure 4: Interaction mode between Q network and environment (2 cranes).

3.3 Parameter optimization in crane scheduling model

To increase the training efficiency of intelligent agent under the framework of deep reinforcement learning, optimization is required in reward discount factor γ , learning rate α , and reward function according to key indexes of convergence rate and training time in the uncertainty task scheduling problem in multi-crane situation.

In order to optimize the parameters in one unified environment, an integrated basic situation of multi-crane scheduling is constructed. Taking five successive transportation tasks as the basic unit of crane scheduling, a

transportation task set consisting of 46 kinds of different scheduling units is created.

- (1) The reward discount factor γ is a factor influencing intelligent agents' attention degree on future reward. The value range is 0–1 and the normal span is 0.5–0.9. This study tests and compares the convergence rates of deep reinforcement learning based crane scheduling model at the γ value of 0.5, 0.6, 0.7, 0.8, and 0.9 for models with different parameters reaching the training goal, the convergence rate is represented by the ratio of completing transportation tasks under the same training step. The comparison result is shown in Figure 5.

It can be seen from Figure 5 that when γ equals 0.8, the training speed is the fastest. Hence, 0.8 is chosen as the value of γ for multi-crane scheduling model.

- (2) Learning rate α determines the magnitude of each parameter updated in Q network model. When the value is relatively low, the model training is stable but slow. When the value is too high, the result can be volatile and cannot converge. Considering the possibility of random coupling between environment status and action, a relatively stable training process is necessary to ensure that the model inclines to converge. Therefore, the value of learning rate α is optimized within the range of 0.0001–0.001. This study tests and compares the convergence rate when α equals to 0.0001, 0.0003, 0.0005, 0.0007, and 0.0009. The result is shown in Figure 6.

It can be seen from Figure 6 that the training speed is fastest when α equals 0.0005. Hence,

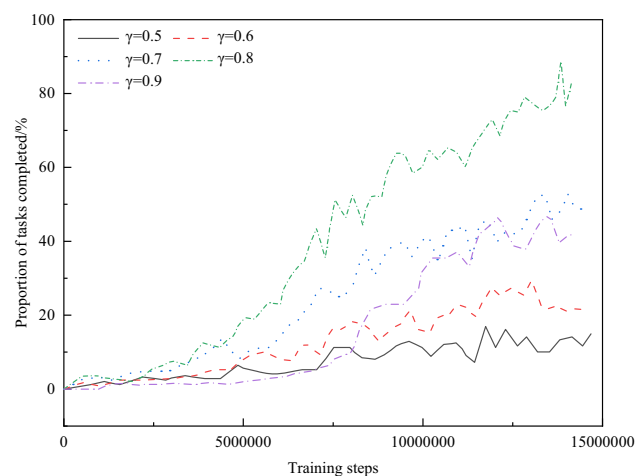


Figure 5: Convergence rate comparison for models with different reward discount factors.

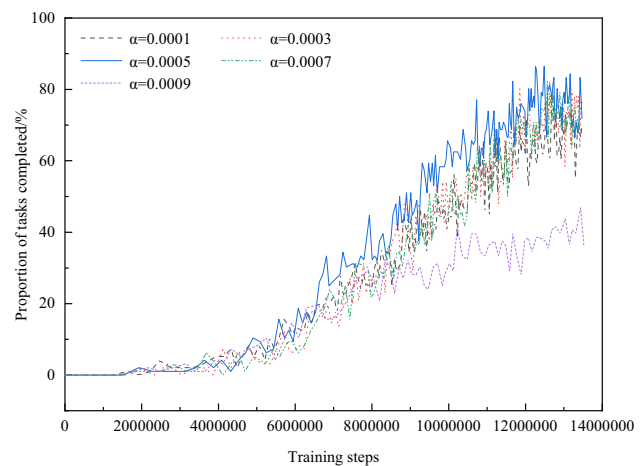


Figure 6: Convergence rate under different learning rates.

0.0005 is adopted as learning rate in multi-crane scheduling model.

- (3) The basic form of reward function is designed as formula (2) in Section 3.2. Intensive improvement is necessary to solve the reward sparsity issue in crane scheduling. Original, weak, and strong reward mechanisms are designed, tested, and analyzed through scheduling process analysis and experts' consultation in different areas. The design of three models are displayed in Tables 1–3.

Based on the original reward mechanism, on the one hand, strong reward mechanism and weak reward mechanism further subdivide the crane status and set different reward values for different crane actions. On the other hand, according to the principle of multiple scaling, weak reward mechanism narrows and strong reward mechanism enlarges the action reward value that involves transportation tasks. Through the setting of three reward mechanisms, the influence of different reward mechanisms on the model training process is tested.

Table 1: Parameters of original reward mechanism

Action	Current status	Reward value
Lifting	At beginning position	5
	Not at beginning position	0
Putting down	At beginning position	30
	Not at beginning position	0
Standstill	Any status	0
Moving along orbit	Any status	0

Table 2: Parameters of weak reward mechanism

Action	Empty crane		Crane loaded with ladle	
	Status	Reward value	Status	Reward value
Lifting	At beginning position	2	Any position	−0.1
	Not at beginning position	0		
Putting down	Any position	−0.1	At target position	5
			Not at target position	−0.1
Standstill	No awaiting task	0.1	Any status	−0.2
	With awaiting task	0		
Moving along orbit	About to collide	−1	About to collide	−1
	About to overpass border	−1	Will not collide and move close to target position	0.1
	Other scenarios	0	Will not collide and move away from target position	−0.2
			About to overpass border	−1

Table 3: Parameters of strong reward mechanism

Action	Empty crane		Crane loaded with ladle	
	Status	Reward value	Status	Reward value
Lifting	At beginning position	5	Any position	−0.1
	Not at beginning position	0		
Putting down	Any position	−0.1	At target position	30
			Not at target position	−0.1
Standstill	No awaiting task	0.1	Any status	−0.2
	With awaiting task	0		
Moving along orbit	About to collide	−1	About to collide	−1
	About to overpass border	−10	Will not collide and move close to target position	0.1
	Other scenarios	0	Will not collide and move away from target position	−0.2
			About to overpass border	−10

With the goal of randomly extracting four groups of basic unit task in crane scheduling, the training time and total task completion time are collected and results displayed in Table 4 under the above three reward mechanisms.

From Table 4 it can be concluded that the training time reduced significantly after intensive improvement in reward mechanism. Considering the total task completion time comprehensively, strong reward mechanism is selected as the training parameter for crane scheduling model.

Table 4: Training effect under different reward mechanisms

Reward mechanism	Training time (h)	Total task completion time (s)
Original	15	1,973
Weak	9	1,933
Strong	10	1,893

4 Experiments and results

4.1 Experiment data

Crane scheduling situations involving both multi-crane scheduling and significant uncertainty mainly occur in industrial production processes. Taking steelmaking process as an example, since there are multiple smelting production lines within one steelmaking workshop, and multiple cranes are shared to complete transportation tasks between the production procedures, it is a typical multi-crane scheduling situation. In the meantime, as the smelting time of production procedure is volatile to some extent, the crane tasks have apparent uncertainty, further increasing the difficulty of crane scheduling.

The crane scheduling process of steel span in an actual steelmaking workshop is selected as the situation to test and verify the proposed model. The layout of the span is shown in Figure 7. The total length of the span is

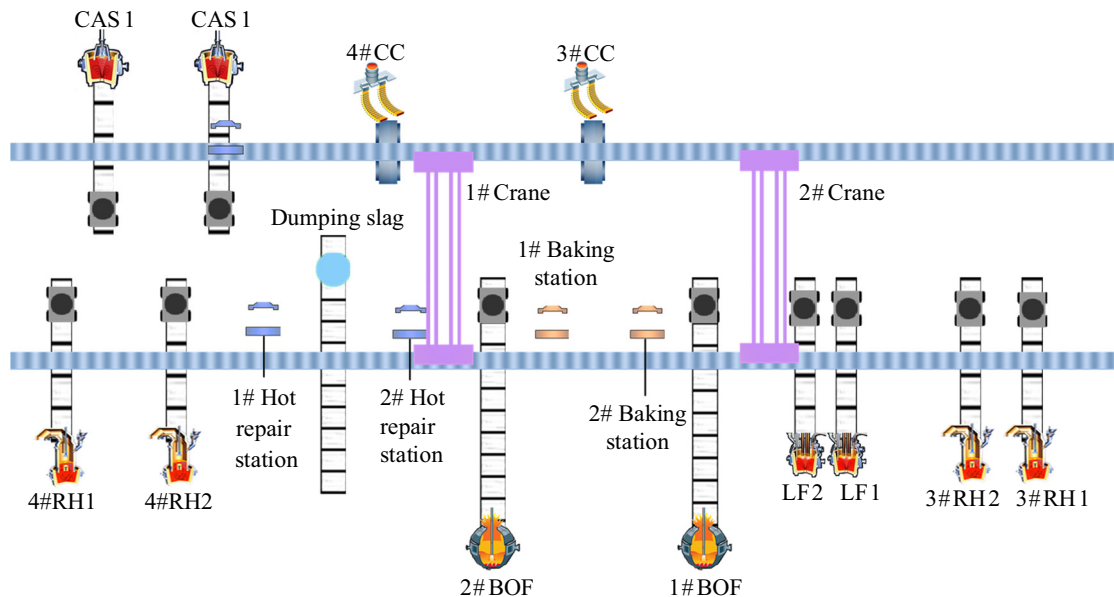


Figure 7: Layout of steel span in one steelwork.

300 m, equipped with 2 converters, 2 double-position RH refining furnaces, 1 double-position LF refining furnaces, 1 double-position CAS refining furnaces, 2 continuous casting machine, 2 baking stations, 2 thermal repair stations, 1 dumping position, and 2 cranes.

The transportation tasks in the span include two phases: heavy ladle (loaded with molten steel) transport and empty ladle (without molten steel) transport. Heavy ladle transport consists of molten steel transported from converter to refinery, from refinery to continuous casting, and some from LF refining to RH refining. Empty ladle transport includes the slag turning activity after continuous

casting, transported from dumping position to thermal repair position, from thermal repair to baking station, and from baking station to converter position.

In total, 10,900 items of crane tasks are collected and filtered according to time sequence, based on the actual production data and crane operation data in the workshop between January and March 2016. Some crane task data are shown in Table 5. Taking the first 10,000 items of data, set every 5 consecutive tasks as one basic unit, and form the basic unit set of training model. Take the latter 900 items of data and randomly intercept every 100 consecutive tasks as the testing set of the model. Then, test

Table 5: List of some crane tasks

Number	Task assigned–target completion time	Beginning position	Target position
1	00:00:30–00:10:10	1#BOF	3#RH1
2	00:15:05–00:23:05	2#BOF	4#RH1
3	00:55:05–01:04:10	3#CC	Slab turning position
4	00:59:15–01:24:10	3#RH1	3#CC
5	01:12:05–01:37:00	4#RH1	4#CC
6	01:46:15–02:10:00	Slab turning position	2#thermal repair
7	02:04:15–02:12:00	1#BOF	LF1
8	02:12:05–02:20:00	4#CC	Slab turning position
9	02:23:10–02:32:00	2#thermal repair	2#BOF
10	02:30:05–02:39:10	3#CC	Slab turning position
11	02:47:05–02:55:00	Slab turning position	2#thermal repair
12	02:59:15–03:19:10	LF1	4#CC
13	03:00:05–03:20:00	2#Thermal repair	1#BOF
14	03:05:05–03:14:10	Slab turning position	1#thermal repair
15	03:09:15–03:17:00	1#BOF	LF1

and verify the executive effect of the crane scheduling plan generated by the model.

4.2 Deep reinforcement learning based crane scheduling plan

This study constructed the deep reinforcement learning-based crane scheduling model for uncertainty tasks under multi-crane situation by applying Python 2.7 language, deep learning framework Pytorch 0.2.0, and reinforcement learning platform OpenAI Gym. The major parameters of the model are learning rate $\alpha = 0.0005$, reward discount factor $\gamma = 0.8$, and playback memory pool size $M = 1$ million, and a strong reward mechanism is adopted as the intensive strategy of reward function.

Based on historic task data, the ratio that the model completes task sequence gradually becomes stable and close to 100% convergence rate after 9 million steps of iterative training. The convergence curve of the training process is displayed in Figure 8.

The crane scheduling plan for testing set is created based on the model and the time-space orbit graphics of the cranes are drawn, some graphics are shown in Figure 9.

From the analysis of the graphics, it can be seen that there is no collision or moving out of the orbit range during the crane operation process. In the meanwhile, when comparing the task completion time in the scheduling plan and the target completion time in the task list, all tasks are completed within the target time. Based on these results, it can be concluded that the scheduling

plan can effectively complete the defined crane tasks and the feasibility of deep reinforcement learning-based crane scheduling method is verified.

4.3 Comparison between deep reinforcement learning-based scheduling plan and manual scheduling plan

Currently there is no crane scheduling model that can be applied in practical production process, due to the uncertainty in crane tasks and the difficulty in dynamic coordination between multiple cranes in steelmaking workshops. In all steelmaking workshops, manual command is applied, allocating cranes to transportation tasks dynamically and coordinating the routes planning between multiple cranes.

To examine the high effectiveness of the deep reinforcement learning-based crane scheduling model proposed by this study, the crane scheduling plan generated by the model is compared with the manual scheduling plan in actual production process (with same 100 items of crane tasks). The advantages and disadvantages of two scheduling plans are analyzed and assessed from three dimensions: task completion time, times of routes collision, and return distance when routes collide. The comparison result is shown in Table 6.

It can be observed that, compared to manual scheduling plan, deep reinforcement learning-based crane scheduling plan shortened the task completion time by 1,405 s or 11.52%, which indicates that the transportation

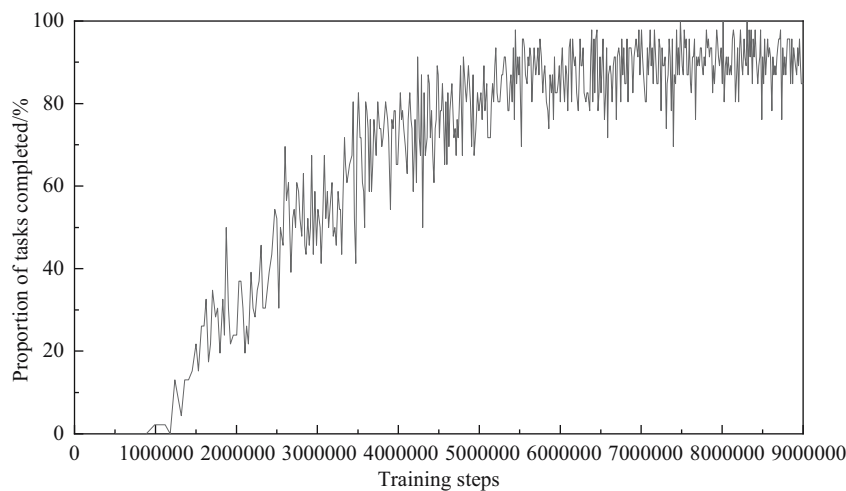


Figure 8: Convergence curve of training process in crane scheduling model.

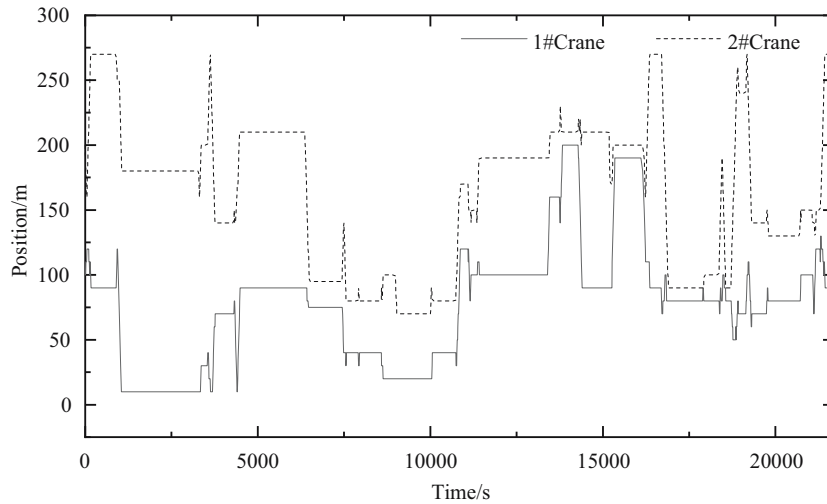


Figure 9: Some time–space orbit graphics of cranes.

Table 6: Comparison results of two scheduling plans

Scheduling plan	Total task completion time (s)	Times of routes collision	Return distance when collide (m)
Manual scheduling plan	12,195	21	760
Deep reinforcement learning-based crane scheduling plan	10,790	9	340

task can be completed within a shorter time. The times of routes collision decreased by 12 or 57.14%, passive crane transportation distance reduced by 420 m or 55.26%, proving that the proposed method has more reasonable route planning and lower risk of crane collision, thereby increasing the efficiency of cranes.

All the above comparisons indicate that, through reasonable design of deep learning network and reinforcement learning mechanism, and optimization of parameter setting, especially intensive improvement in reward function, the deep reinforcement learning-based crane scheduling model can quickly grasp the scheduling strategy of completing task and reducing route collision, thereby generating a more effective and efficient crane scheduling plan.

Compared with manual scheduling method, crane scheduling method based on deep reinforcement learning can explore potential data patterns through iterative training of transportation task data-set, and form a more reasonable and efficient scheduling plan than human experience. The main reason is that, on the one hand, the transportation task data-set combined with actual production data contains more comprehensive task types and composition relationships, and its amount of information far exceeds the memory and experience of

dispatchers. On the other hand, the method based on deep reinforcement learning can approach the global optimum scheduling plan through continuous iterative training. The space span, time span, and task volume considered by the scheduling plan far exceed the capability of the dispatchers.

5 Conclusion

Aiming at the crane scheduling problem for uncertainty tasks in multi-crane scheduling situation, this article proposes a deep reinforcement learning-based crane scheduling model. By testing with practical production data in crane scheduling situation in steelworks, the scheduling plan based on the model is proved to be effective and efficient.

The method is based on the deep reinforcement learning framework, with the orbit space of the crane and the transportation task as environmental information and crane as the agent. First, reasonable status space, action space, and reward function are designed. Second, the interactive mode between reinforcement learning algorithm and the environment is adjusted to adapt to

the combined training of multi-crane scheduling model. Last, the reward discount factor, learning rate, and reward function intensive mode are optimized. The crane scheduling model applicable for uncertainty tasks is then constructed.

Training and testing of the model are carried out with actual crane tasks in steelworks. Results show that all tasks are completed within the planned period in the deep reinforcement learning-based scheduling plan, proving its feasibility. Compared with manual scheduling plan, the scheduling proposal based on the new model reduces the total task completion time by 1,405 s or 11.52%. Times of collision of crane route decrease by 12 or 57.14%. Negative crane transportation distance shortens by 420 m or 55.26%. The high effectiveness and efficiency of the scheduling model is therefore verified.

The crane scheduling method based on deep reinforcement learning has strong application potential in crane scheduling optimization of production workshops in iron and steel enterprises, considering its ability to carry out model training based on large-scale transportation task data without relying on human experience or setting of objective functions, and can approach the global optimum scheduling plan through continuous iteration. In future research and application, not only multiple cross-region joint scheduling optimization in production workshop can be attempted, but also dynamically update scheduling plan and strategy through parallel training with production process, in order to adapt timely to the evolution of production rhythm and working conditions.

Acknowledgment: In addition to the support from relevant funds, the research process of this paper also received the help of Professor Xu An-jun's team, Professor Zhang De-zheng's team and Dr. Xu Cong. I would like to express my sincere thanks.

Funding information: This research was funded by National Key Research and Development Program (No. 2017YFB0304001), National Natural Science Foundation of China (No. 51674030), and Fundamental Research Funds for the Central Universities (No. FRF-TP-16-081A1).

Author contributions: Kai Feng was responsible for the design of research ideas and system framework; Lingzhi Yang provided technical support and problem solving; Shijing Lin was responsible for the specific programming work; Dongfeng He and Buxin Su were responsible for the testing of models and data.

Conflict of interest: The authors state no conflict of interest.

Data Availability Statement: Limited by cooperation with enterprises, this article does not provide publicly available data.

References

- [1] Kim, K. H. and Y. M. Park. A crane scheduling method for port container terminals. *European Journal of Operational Research*, Vol. 156, No. 3, 2004, pp. 752–768.
- [2] He, J., Y. Huang, and W. Yan. Yard crane scheduling in a container terminal for the trade-off between efficiency and energy consumption. *Advanced Engineering Informatics*, Vol. 29, No. 1, 2015, pp. 59–75.
- [3] Zhao, G., J. Liu, and Y. Dong. Scheduling the operations of a double-load crane in slab yards. *International Journal of Production Research*, Vol. 58, No. 9, 2020, pp. 2647–2657.
- [4] Xie, X., Z. Li, and Z. Yongyue. A Hybrid Sub-Heuristic Algorithm for Solving a Single Crane Scheduling Problem with Heat Loss. *Journal of Shenyang University (Natural Science)*, Vol. 31, No. 2, 2019, pp. 107–112.
- [5] Tang, L., X. Xie, and J. Liu. Crane scheduling in a warehouse storing steel coils. *lie Transactions*, Vol. 46, No. 3, 2014, pp. 267–282.
- [6] Ng, W. C. and K. L. Mak. An effective heuristic for scheduling a yard crane to handle jobs with different ready times. *Engineering Optimization*, Vol. 37, No. 8, 2005, pp. 867–877.
- [7] He, J., H. Yu, C. Tan, W. Yan, and C. Jiang. Quay crane scheduling for multiple hatches vessel considering double-cycling strategy. *Industrial Management & Data Systems*, Vol. 120, No. 2, 2019, pp. 253–264.
- [8] Ma, S., H. Li, N. Zhu, and C. Fu. Stochastic programming approach for unidirectional quay crane scheduling problem with uncertainty. *Journal of Scheduling*, Vol. 24, No. 2, 2021, pp. 137–174.
- [9] Heshmati, S., T. A. M. Toffolo, W. Vancroonenburg, and G. V. Berghe. Crane-operated warehouses: Integrating location assignment and crane scheduling. *Computers & Industrial Engineering*, Vol. 129, 2019, pp. 274–295.
- [10] Wang, X., S.-X. Liu, and J. Wang. Memetic algorithm for crane scheduling problem in slab yard with spatial and temporal constraints. *Journal of Northeastern University Nature Science*, Vol. 38, No. 7, 2017, pp. 913–917.
- [11] Chengji, L. and L. Yang. Research on problem of double-trolley quay crane and AGV coordinated scheduling in automated terminal. *Computer Engineering and Applications*, Vol. 55, No. 10, 2019, pp. 256–263.
- [12] Hu, H., X. Chen, L. Zhen, C. Ma, and X. Zhang. The Joint quay crane scheduling and block allocation problem in container terminals. *IMA Journal of Management Mathematics*, Vol. 30, No. 1, 2019, pp. 51–75.
- [13] Xie, X., Y. Zheng, and Y. Li. Multi-crane scheduling in steel coil warehouse. *Expert Systems with Applications*, Vol. 41, No. 6, 2014, pp. 2874–2885.
- [14] Cheng, X., L. Tang, and P. M. Pardalos. A Branch-and-Cut algorithm for factory crane scheduling problem. *Journal of Global Optimization*, Vol. 63, No. 4, 2015, pp. 729–755.

- [15] Li, J., A. Xu, and X. Zang. Simulation-based solution for a dynamic multi-crane-scheduling problem in a steelmaking shop. *International Journal of Production Research*, Vol. 9, 2019, pp. 1–15.
- [16] Zheng, Z., C. Zhou, and K. Chen. Crane scheduling simulation model based on immune genetic algorithms. *Systems Engineering-Theory & Practice*, Vol. 33, No. 1, 2013, pp. 223–229.
- [17] Sun, L., W. Liu, T. Chai, H. Wang, and B. Zheng. Crane scheduling of steel-making and continuous casting process using the mixed-timed petri net modelling via CPLEX optimization. *IFAC Proceedings Volumes*, Vol. 44, No. 1, 2011, pp. 9482–9487.
- [18] Han, X. L., Z. Q. Lu, and L. F. Xi. A proactive approach for simultaneous berth and quay crane scheduling problem with stochastic handling time. *European Journal of Operational Research*, Vol. 207, No. 3, 2009, pp. 1327–1340.
- [19] Daganzo, C. F. The crane scheduling problem. *Transportation Research Part B*, Vol. 23, No. 3, 2008, pp. 159–175.
- [20] Lim, A., B. Rodrigues, and Z. Xu. A m-parallel crane scheduling problem with a non-crossing constraint. *Naval Research Logs*, Vol. 54, No. 2, 2010, pp. 115–127.
- [21] Lei, D., P. Zhang, Y. Zhang, Y. Xia, and S. Zhao. Research on optimization of multi stage yard crane scheduling based on genetic algorithm. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 2, 2020, pp. 483–494.
- [22] Zheng, F., X. Man, F. Chu, M. Liu, and C. Chu. A two-stage stochastic programming for single yard crane scheduling with uncertain release times of retrieval tasks. *International Journal of Production Research*, Vol. 57, No. 13–14, 2019, pp. 4132–4147.
- [23] Tanizaki, T., H. Katagiri, and A. O. N. René. Scheduling algorithms using metaheuristics for production processes with crane interference. *International Journal of Automation Technology*, Vol. 12, No. 3, 2018, pp. 297–307.
- [24] Kasm, O. A., A. Diabat, and T. C. E. Cheng. The integrated berth allocation, quay crane assignment and scheduling problem: mathematical formulations and a case study. *Annals of Operations Research*, Vol. 291, 2020, pp. 435–461.
- [25] Dohn, A. and J. Clausen. Optimising the slab yard planning and crane scheduling problem using a two-stage heuristic. *International Journal of Production Research*, Vol. 48, No. 15, 2010, pp. 4585–4608.
- [26] Arai, S., K. Miyazaki, and S. Kobayashi. Controlling multiple cranes using multi-agent reinforcement learning: Emerging coordination among competitive agents. *IEICE Transactions on Communications*, Vol. E83B, No. 5, 2000, pp. 1039–1047.
- [27] Cui, R., C. Yang, Y. Li, and S. Sharma. Adaptive neural network control of AUVs with control input nonlinearities using reinforcement learning. *IEEE Transactions on Systems Man & Cybernetics Systems*, Vol. 47, No. 6, 2017, pp. 1019–1029.
- [28] Findi, A. H. M., M. H. Marhaban, R. Kamil, and M. K. Hassan. Collision prediction based genetic network programming-reinforcement learning for mobile robot navigation in unknown dynamic environments. *Journal of Electrical Engineering & Technology*, Vol. 12, No. 2, 2017, pp. 890–903.
- [29] Mannion, P., J. Duggan, and E. Howley. Parallel reinforcement learning for traffic signal control. *Procedia Computer Science*, Vol. 52, 2015, pp. 956–961.
- [30] Arulkumaran, K., M. P. Deisenroth, M. Brundage, and A. A. Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, Vol. 34, No. 6, 2017, pp. 26–38.
- [31] Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, et al. Human-level control through deep reinforcement learning. *Nature*, Vol. 518, No. 7540, 2015, pp. 529–533.
- [32] Mnih, V., K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, et al. Playing Atari with deep reinforcement learning. *Computer Science*, 2013. <http://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>.
- [33] Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, et al. Mastering the game of Go without human knowledge. *Nature*, Vol. 550, No. 7676, 2017, pp. 354–359.