Alexandria Engineering Journal (2023) 67, 397–407

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Deep reinforcement learning based optimization of automated guided vehicle time and energy consumption in a container terminal

Darius Drungilas [a,*], Mindaugas Kurmis [a], Audrius Senulis [a], Zydrunas Lukosius [a], Arunas Andziulis [a], Jolanta Januteniene [a], Marijonas Bogdevicius [a], Valdas Jankunas [a], Miroslav Voznak [a,b]

[a] *Klaipeda University, H. Manto str. 84, LT-92294 Klaipeda, Lithuania*
[b] *Dept. of Telecommunications, VSB - Technical University of Ostrava, 17. Listopadu 2172/15, 708 00 Ostrava, Czech Republic*

Received 10 June 2022; revised 21 December 2022; accepted 26 December 2022

KEYWORDS

Automated guided vehicle (AGV);
Container terminal;
Energy consumption;
Deep reinforcement learning;
Modeling;
Optimization

Abstract The energy efficiency of port container terminal equipment and the reduction of CO_2 emissions are among one of the biggest challenges facing every seaport in the world. The article presents the modeling of the container transportation process in a terminal from the quay crane to the stack using battery-powered Automated Guided Vehicle (AGV) to estimate the energy consumption parameters. An AGV speed control algorithm based on Deep Reinforcement Learning (DRL) is proposed to optimize the energy consumption of container transportation. The results obtained and compared with real transportation measurements showed that the proposed DRL-based approach dynamically changing the driving speed of the AGV reduces energy consumption by 4.6%. The obtained results of the research provide the prerequisites for further research in order to find optimal strategies for autonomous vehicle movement including context awareness and information sharing with other vehicles in the terminal.

1. Introduction

The European Environment Agency (EEA) (2020) data show that there has been a steady overall reduction in greenhouse gas (GHG) emissions in the EU in recent years. However, the transport sector has not followed this general trend and, as a result, its relative contribution to overall GHG emissions in Europe have become more significant. According to Europe report [1] of Greenhouse gas emissions from transport, without the implementation of additional measures, an increase could be observed until 2025, and the reduction expected thereafter would still leave transport emissions in 2030 around 10% above 1990 levels. The transport sector CO_2 emissions represent more than one-fourth of total EU28 greenhouse gas (GHG) emissions in 2017.

* Corresponding author.
E-mail address: darius.drungilas@ku.lt (D. Drungilas).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.12.057
1110-0168

In recent years, the research about the optimization of energy consumption has been spreading over the world. Some studies [2] suggest that energy innovation helps to reduce carbon emissions however others express their concerns about the negative environmental impacts for technological progress.

According to the results of the investigations of CO_2 emissions in port container terminals [3,4] the yard terminal tractors, rubber tyred gantry cranes (RTGs) and quay cranes are the main emission sources. This leads the researchers and terminal operators to focus on the research of energy efficiency in ports terminals. Iris and Lam in [5] has made one of the first comprehensive literature reviews of operational strategies, technologies, and energy management systems for energy efficiency in ports. The energy savings and emission reductions can be achieved with energy management, state-of-the-art technologies, and operational improvements. Researchers state that some ports have phased in electrified/hybrid equipment such as E-RTG, BAGVs, ALVs, IAVs. There is no single method, technology or management system that dominates the remainder with respect to energy efficiency, cost of investment and ease of implementation. According to Yang Y.C. [6] the optimal green port assessment criteria could lead to lessening the berthing time of ships in port and reduce energy costs and CO_2 emissions. Another case study of the Shekou Container Terminal evaluating the emission reduction strategies for berthing containerships concluded that a combination of shore power system and increasing quay crane efficiency has the highest emission reduction potential [7].

The solution to increase an efficiency, save energy and reduce emissions in container terminals is to optimize and modernize current equipment and processes and the most accessible method is to optimize the yard crane and automated guided vehicle (AGV) scheduling and path planning operations. For instance, Chen et al. [8] developed an efficient routing for multi-AGV by combining centralized control with decentralized control, based on the Ant-agent and this allowded to improve the performance on collision avoidance, transportation distance and transportation efficiency. In another similar study [9] a 10.98% reduction in AGV energy consumption was achieved by applying energy benchmark methodology for energy-efficient path planning. Yu et al. [10] have focused on the emissions from yard tractors during loading in container terminals. The bilevel bi-objective model for AGV scheduling problem enabling reduction of energy consumption has been designed by Wang et al. [11]. The numerical experiments have verified the effectiveness of the bilevel model as the overall energy consumption decreases by 11.65%. The yard crane scheduling problem has also been analyzed by He et al. [12] where authors formulated a mixed-integer programming (MIP) model with two objectives: to minimize the total completion delay of all task groups and the total energy consumption of all YCs. The authors have developed an integrated simulation optimization method with genetic algorithm (GA) and the particle swarm optimization (PSO) algorithms. The simulation results showed that the proposed method can solve YC scheduling problems with varied sizes however in real-world the arriving time and handling volume of each task group would fluctuate and cannot be predicted accurately by the proposed model. One of the most promising methods to solve scheduling problems is reinforced learning methods. The flow-shop scheduling problem (FSP) is one of the most widely studied job-shop scheduling problems (JSPs) and could be applied to many different engineering areas. Ren J. et al. [13] was solving flow-shop scheduling problem with a reinforcement learning algorithm that generalizes the value function with neural network. Experimental results showed that the proposed algorithm framework can effectively solve FSPs however additional research is needed to be capable of solving JSPs of various scales.

With the development of AI technologies, more and more studies have attempted to apply deep learning and reinforcement learning methods for energy efficiency improvement. For instance, the renewable power source energy consumption by hybrid machine learning model by using CatBoost, Multilayer Perception and Support Vector Regression methods has been developed by [14]. [15] has developed microgrid energy management using deep Q-network reinforcement learning. Also, [16] considers energy management of intelligent building based on deep reinforced learning. Others [17] analyze reliability-driven distribution power network dynamic reconfiguration by the deep reinforcement learning methods.

Authors in [18] have proposed a DRL approach that uses a deep deterministic policy gradient algorithm for integrated control of HVAC and electric battery storage systems in the presence of on-site PV generation. The DRL algorithm has been trained on synthetic data and were evaluated on the physical test building. The performance in delivering energy efficiency, load shift, and load shed was evaluated using price-based signals. The results showed that the DRL-based controller can produce cost savings of up to 39.6% as compared to the baseline controller while maintaining similar thermal comfort in the building.

Ren et al. [19] proposed a deep neural network-based real-time optimal navigation for an automatic guided vehicle with static and dynamic obstacles. The performed numerical experiments demonstrated that the proposed model was able to generate the optimal control instructions on-board to steer the AGV to the desired location under the ideal simulated conditions.

Another approach to minimize carbon dioxide emissions during the container handling at marine container terminals was made by using the hybrid evolutionary algorithms [20]. The paper presents a mixed-integer mathematical model for the berth scheduling problem, which minimizes the total service cost of vessels, including the total carbon dioxide emission cost due to container handling. Unlike typical stochastic search algorithms, the proposed algorithm deployed a set of local search heuristics to facilitate exploration and exploitation of the search space.

A new method for marshaling plan using a reinforcement learning considering the desired layout of containers in port has been developed by Hirashima et al. [21]. In the proposed method, the learning process consists of two parts: rearrangement plan assuring explicit transfer of container to the desired position, and removal plan for preparing the rearrange operation. By using simulation authors have proved that the learning performance can be improved as compared to the conventional methods.

Eglynas et al. [22] made research and evaluation of the energy consumption of container diesel trucks in a container terminal at Klaipeda port. Authors developed a mathematical model which describes the instantaneous energy consumption of the diesel trucks, considering their dynamic properties and the overall geometry of their routes "Ship-Truck-Stack-Ship" using the superposition principle. The modeling results showed that an instantaneous evaluation of energy consumption can reveal areas in the container transportation process that have the highest energy loss and require the introduction of new management and process control initiatives.

However, no papers were found on the application of deep reinforcement learning for optimization of energy consumption and reduction of emissions in the AGVs.

To develop an integrated energy-efficient adaptive autonomous crane and autonomous electric vehicle control system, there is a contradiction between energy efficiency and process time. The increase of the loading time reduces the energy consumption. To reduce the loading and transferring time, it is necessary to increase energy consumption. This optimization task is also influenced by the mass of the cargo, the maximum allowed acceleration of the container, traffic in the container terminal, and the weather conditions. It is necessary to determine the influence of these parameters on the loading process, to find the total minimum energy consumption under the appropriate conditions.

In this paper, we propose a new deep reinforcement learning-based AGV control approach which will be able to optimize the energy consumption of the AGV by dynamically controlling movement speed according to the route conditions at container terminal territory.

2. Deep reinforcement learning for AGV control

The organization of AGV work in port container terminals is very important, affecting the speed of container transportation and pollution of the port environment, which is directly proportional to energy consumption. Container transportation in the terminal area from the quay crane to the stack and vice versa in many terminals is still performed by trucks operated by a human. This study is based on the hypothesis that the transportation of containers by a human-operated truck at the container terminal is not optimal and energy consumption or transportation time can be optimized. Usually, when transporting a container in the terminal area, AGV moves on fixed trajectories that have certain speed sections. From a modeling point of view, a speed limit must be set at a certain point on the trajectory. Therefore, the AGV speed controller must comply with these restrictions and at the same time find the optimal balance between energy consumption and transport time.

Finding an optimal strategy for transporting the container from the quay crane to the stack in terms of energy consumption can be formulated as a reinforcement learning control problem where the environment is formulated as Markov

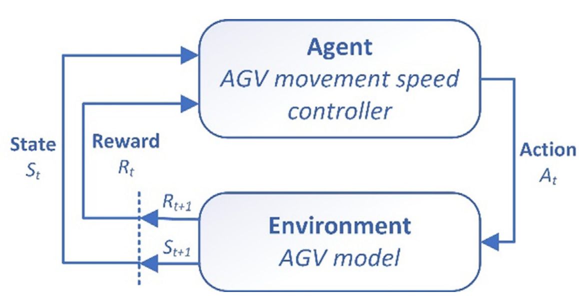

Fig. 1 Reinforcement learning control problem as MDP [23].

Decision Process (MDP) (Fig. 1). According to the standard reinforcement learning setup, an agent interacts with an environment in discrete time steps by receiving observation and applying an action, and trying to find an optimal policy, when the agent receives a positive or a negative reward while performing an action.

In the case of this study, the environment is represented by the AGV model, where observation parameters (state S) – instantaneous power, movement speed, and acceleration are measured and target movement speed actions A from continuous spaces are applied. According to the types and dimensions of the environment input and output parameters, an actor-critic, model-free algorithm based on the deep deterministic policy gradient (DDPG) [24] is used, that can solve control problems where large dimensional and continuous state and action spaces are required.

2.1. Model of environment

In this research, the environment interacted by the DRL agent is represented by the AGV model which takes action to set the speed of movement and outputs the state parameters at a fixed time step.

2.1.1. AGV motion model

The AGV movement consists of several components:

1. A model of resistance forces to AGV motion, describing the forces acting on AGV motion. The following assumptions and modelling parameters were applied to this model:
   a. The territory of the port container loading terminal has a flat surface, therefore the forces influencing the movement uphill or downhill are not evaluated [22,25];
   b. AGV moves in the port area where there are many high elements - stacks, ship, warehouses, as well as AGV moves quite slowly (up to 8.3 m/s) [22,26], therefore only air resistance is assessed without assessing wind directions;
   c. The condition of the paved road surface - ice, snow, wet or dry - has not been analyzed with a view to addressing it in future studies.

2. AGV electric drive model. The following assumptions and modelling parameters were applied to this model:
   a. AGV power source - $LiFePO_4$ batteries [27];
   b. AGV has two asynchronous motors with increased starting torque [28,29];
   c. The gear unit is evaluated by the gear ratio ($i = 10$) and the efficiency factor ($\eta_r = 0.9$) [30], without modeling the mechanical drive in detail;
   d. Wheel diameter – 1 m;
   e. DC / AC converter - adopted Matlab/Simulink model consisting of three-phase MOSFET inverter [31];
   f. Drive control system - a slave consisting of a speed PI controller and a current two-position controller and a PWM signal generator that controls the DC / AC converter [31].

Thus, the AGV motion resistance force model hereinafter referred to as the load model, consists of two main modes:

1. When the AGV is standing;
2. When the AGV is driving:
   a. When the AGV is driving with acceleration (positive and negative) on a flat road;
   b. When the AGV is traveling at a constant speed on a flat road.

The resistance forces of the AGV motion are defined as follows:

1. Rolling friction $F_{fr.}$:

$$F_{fr.} = mgf \tag{1}$$

Were m is the AGV mass, g is the free fall acceleration, f is the rolling friction coefficient, which depends on: $f = f_0 \cdot (1 + A_0 \cdot v^2)$, f_0 – is the coefficient of friction set at low speeds between 0.006 and 0.01 [32] between the rubber tire and the asphalt; A_0 - empirical coefficient – (4.5) 10^{-5}.

2.Air resistance force F_{air}:

$$F_{air} = \frac{1}{2}\rho_{air}C_d A v^2 \tag{2}$$

where ρ_{air} - air density, C_d - aerodynamic fluidity coefficient for AGV it is from 0.6 to 0.7 [20], $A = h_a{\cdot}b$ area of AGV front, h_a - AGV height, b - AGV width, v - AGV speed.

3.Force of inertia resistance F_{in}:

$$F_{in} = F_{in1} + F_{in2} \tag{3}$$

where F_{in1} is the inertia force estimating the mass of the AGV, F_{in2} is the inertia estimating the inertia of the rotating parts (wheels, motors, reducers). While accelerating the force is positive, decelerating - force is negative.

The above forces of inertia F_{in1} and F_{in2}:

$$F_{in1} = ma' \tag{4}$$

$$F_{in2} = J\frac{i^2}{r^2}a \tag{5}$$

where J is the equivalent moment of inertia of all rotating parts, i is the gear ratio, r is the AGV wheel radius, a' is the AGV acceleration.

4.Total force F of resistance to movement:

$$F = F_{fr.} + F_{air} + F_{in} = mgf_0(1 + A_0v^2) + \frac{1}{2}\rho_{air}C_d A v^2 \pm \left(m + J\frac{i^2}{r^2}\right)a' \tag{6}$$

5.Total moment M of resistance to movement:

$$M = \frac{r}{i\eta_r}\left[mgf_0(1 + A_0v^2) + \frac{1}{2}\rho_{air}C_d A v^2 \pm \left(m + J\frac{i^2}{r^2}\right)a'\right] \tag{7}$$

6.Total power P of resistance to movement:

$$P = v\left[mgf_0(1 + A_0v^2) + \frac{1}{2}\rho_{air}C_d A v^2 \pm \left(m + J\frac{i^2}{r^2}\right)a'\right] \tag{8}$$

2.1.2. AGV electric drive model

The AGV electric drive consists of two asynchronous electric motors, with two cages, and their models are described by the equations below. The motor model was changed from a three-phase to a two-phase model using the abc-dq transformation. Motor voltage balance equations [31]:

$$V_{qs} = R_s i_{qs} + \frac{d\varphi_{qs}}{dt} + \omega\varphi_{ds} \tag{9}$$

$$V_{ds} = R_s i_{ds} + \frac{d\varphi_{ds}}{dt} - \omega\varphi_{qs} \tag{10}$$

$$0 = R'_{r1}i'_{qr1} + \frac{d\varphi'_{qr1}}{dt} + (\omega - \omega_r)\varphi'_{dr1} \tag{11}$$

$$0 = R'_{r1}i'_{dr1} + \frac{d\varphi'_{dr1}}{dt} - (\omega - \omega_r)\varphi'_{qr1} \tag{12}$$

$$0 = R'_{r2}i'_{qr2} + \frac{d\varphi'_{qr2}}{dt} + (\omega - \omega_r)\varphi'_{dr2} \tag{13}$$

$$0 = R'_{r2}i'_{dr2} + \frac{d\varphi'_{dr2}}{dt} - (\omega - \omega_r)\varphi'_{qr2} \tag{14}$$

Motor qd system flow equations for modeling [31]:

$$\varphi_{qs} = L_s i_{qs} + L_m\left(i'_{qr1} + i'_{qr2}\right) \tag{15}$$

$$\varphi_{ds} = L_s i_{ds} + L_m\left(i'_{dr1} + i'_{dr2}\right) \tag{16}$$

$$\varphi'_{qr1} = L'_{r1}i'_{qr1} + L_m i_{qs} \tag{17}$$

$$\varphi'_{dr1} = L'_{r1}i'_{dr1} + L_m i_{ds} \tag{18}$$

$$\varphi'_{qr2} = L'_{r2}i'_{qr2} + L_m i_{qs} \tag{19}$$

$$\varphi'_{dr2} = L'_{r2}i'_{dr2} + L_m i_{ds} \tag{20}$$

Motor electromagnetic torque [31]:

$$M_e = 1.5p\left(\varphi_{ds}i_{qs} - \varphi_{qs}i_{ds}\right) \tag{21}$$

Motor mechanical system model [31]:

$$\frac{d\omega_v}{dt} = \frac{1}{2J}\left[M_e - B\omega_v - \mathrm{M}_v\right] \tag{22}$$

$$\frac{d\theta_v}{dt} = \omega_v \tag{23}$$

Where $L_s = L_{ls} + L_m$, $L'_{r1} = L'_{lr1} + L_m$, $L'_{r2} = L'_{lr2} + L_m$, R_s and L_{ls} – stator resistance and leakage inductance, L_m – magnetizing inductance, L_s – total stator inductance, V_{qs} and i_{qs} – q axis stator voltage and current, V_{ds} and i_{ds} – d axis stator voltage and current, φ_{qs} and φ_{ds} – stator q and d axis fluxes, ω_v – angular velocity of the rotor, θ_v – rotor angular position, p – number of pole pairs, ω_r – electrical angular velocity ($\omega_v p$), θ_r – electrical rotor angular position ($\theta_v p$), M_e – electromagnetic torque, M_v – shaft mechanical torque, B – combined rotor and load viscous friction coefficient, R'_{r1} and L'_{lr1} – rotor resistance and leakage inductance of cage 1, R'_{r2} and L'_{lr2} – rotor resistance and leakage inductance of cage 2, L'_{r1} and L'_{r2} – total rotor inductances of cage 1 and 2, i'_{qr1} and i'_{qr2} – q axis rotor current of cage 1 and 2, i'_{dr1} and i'_{dr2} – d axis rotor current of cage 1 and 2, φ'_{qr1} and φ'_{dr1} – q and d axis rotor fluxes of cage 1, φ'_{qr2} and φ'_{dr2} – q and d axis rotor fluxes of cage 2.

Because the model of the drive is created using the Matlab/Simulink block, the model of DC/AC inverter was selected as a universal bridge block, which physically switches and allows using the forced-commutated device. In our model composition, the MOSFET three-phase bridge inverter was selected. The Inverter is commutated using the PWM pulse modulator, which incorporates a cascade control system of speed and motor current.

The battery of the model is modeled by using a standard Matlab/Simulink model with mature lithium-ion battery chemistry of $LiFePO_4$, which is created by using these equations [33–36]:

1. Discharge model (i* > 0):

$$f_1(it, i^*, i) = E_0 - K\frac{Q}{Q-it}i^* - K\frac{Q}{Q-it}it + Ce^{-D\cdot it} \tag{24}$$

2. Charge model (i^* < 0) – at this stage the charge was not simulated in the study that is presented in this paper.

In the equations: E_0 is constant voltage, in V, K is polarization constant, in V/Ah, or polarization resistance, in Ohms, i^* is low-frequency current dynamics, in A, i is battery current, in A, it is extracted capacity, in Ah, Q is maximum battery capacity, in Ah, C is exponential voltage, in V, D is exponential capacity, in Ah^{-1}. During the modeling, the aging and temperature effects were not included in the model.

The control system as mentioned consists of speed feedback, AC feedback, ABC-dq, and dq-ABC conversion, thus a PI speed controller and two-position current regulator, which generates the pulses for the MOSFET converter of the AGV drive. The control system is presented in Fig. 2, which was adapted to the needs of the analyzed AGV system.

Thus, concluding the final AGV drive system is presented in Fig. 3.

Using AGV drive system model, the following parameters are used in observations and reward generation for DRL agent: AGV movement speed [m/s], movement acceleration [m/s^2], power [W] distance [m], time [s].

2.2. Actor-critic model

According to [24] the DDPG agent is designed according to actor-critic architecture, where two deep neural networks are used (Fig. 4). The actor network approximates the policy function by taking observations of the environment and returning the corresponding action that maximizes the long-term reward. The critic network approximates the Q-value function by taking observation and action as inputs and returning the corresponding expectation of the long-term reward. This Q-value approximation method is chosen due to the fact that continuous values (movement speed and acceleration) are used to estimate the state of the AGV, resulting a large state space and, in this case of reinforcement learning, a tabular Q value estimation method is not suitable.

The observation $s_t = [v_t, a'_t, e_{vt}, \Delta e_{vt}]$ at time step t is used by actor and critic where:

v_t – AGV movement speed [m/s], at time step t;

a'_t – AGV movement acceleration [m/s^2], at time step t;

e_{vt} – AGV movement speed error between speed limits at an appropriate position of the route and actual speed [m/s], at time step t;

Δe_{vt} – a derivative of AGV movement speed error between the speed limit and actual speed [m/s], at time step t.

The action is represented by a unary parameter a_t that defines the targe speed for AGV at time step t.

The Q (s, a) function is learned using the Bellman equation [24]:

$$Q^\pi(s_t, a_t) = \mathbb{E}_{s_{t+1}\sim E, a_t\sim\pi}[r(s_t, a_t) + \gamma\mathbb{E}_\pi[Q^\pi(s_{t+1}, a_{t+1})]] \tag{25}$$

where $\gamma \in [0, 1]$ is a discount factor. And an actor policy is updated using the sampled gradient:

$$\nabla_{\theta^\mu}\mu|_{s_t} \approx \frac{1}{N}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i,\ a=\mu(s_i)}\nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i} \tag{26}$$

where $\mu(s|\theta^\mu)$ is a parameterized actor function which ecifies the current policy by deterministically mapping states to a specific action [24].

2.3. Reward estimation

In this research, the agent gets a total reward by summing up penalties and rewards of appropriate parameters of the envi-

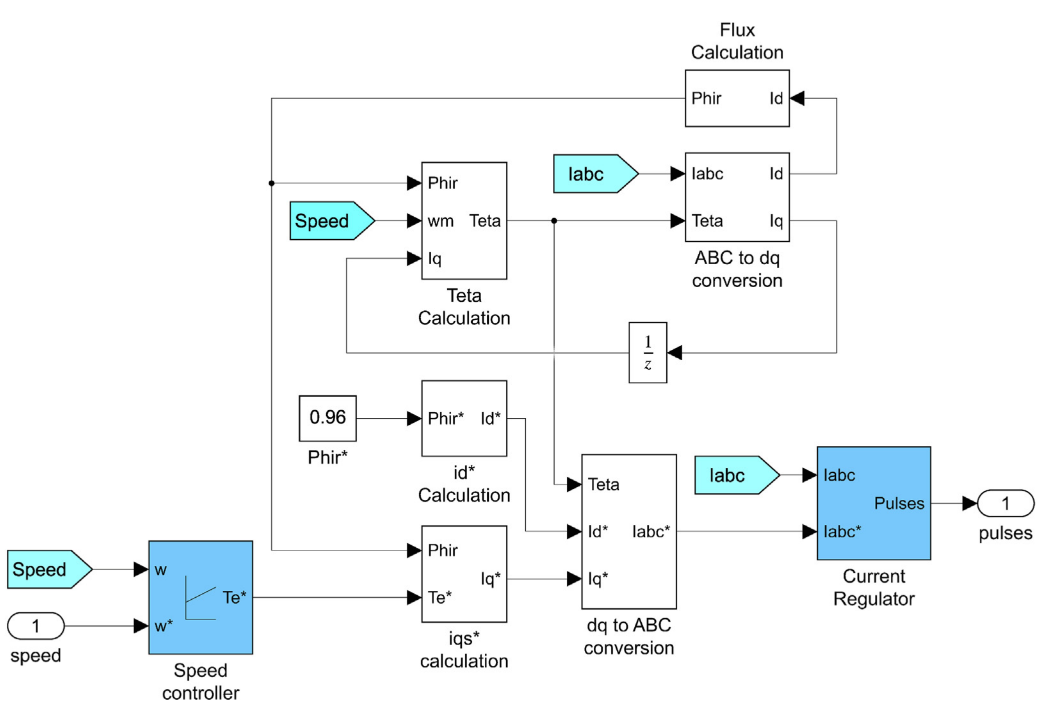

Fig. 2 Cascade Speed, current control system of AGV drive.

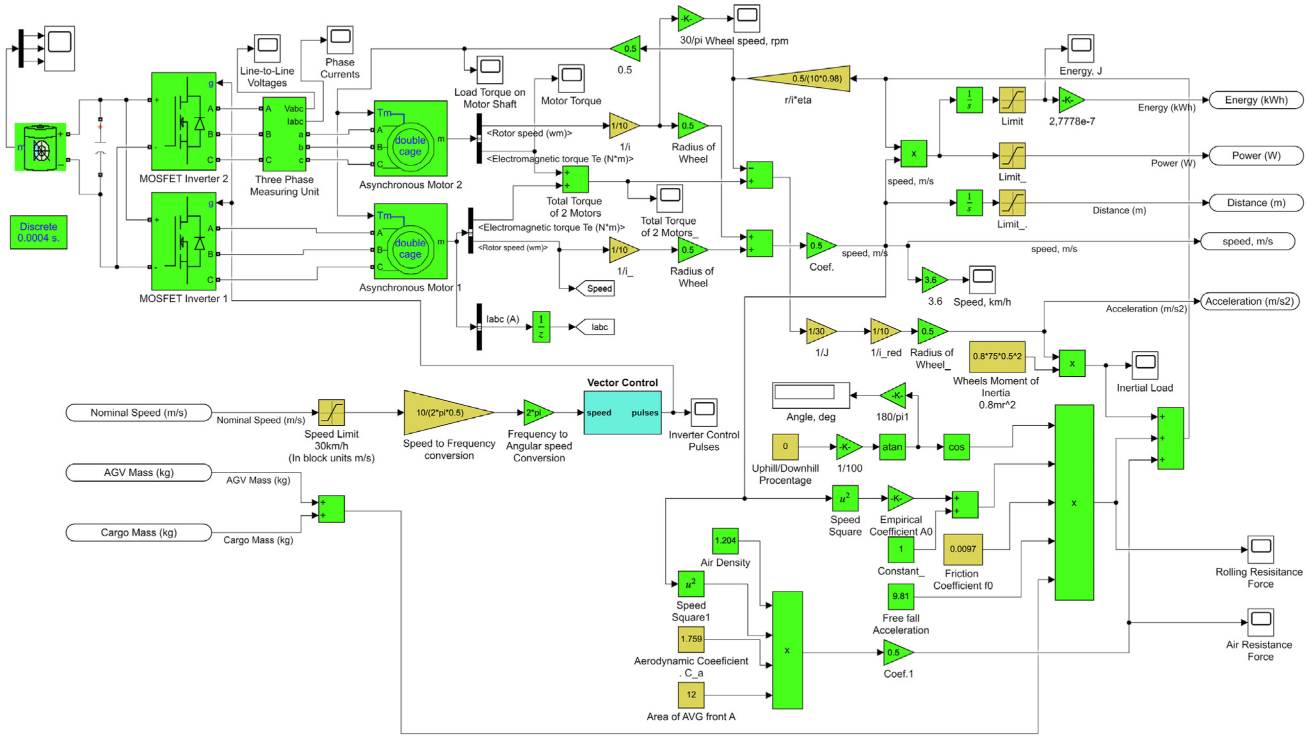

Fig. 3 AGV drive system Simulink model.

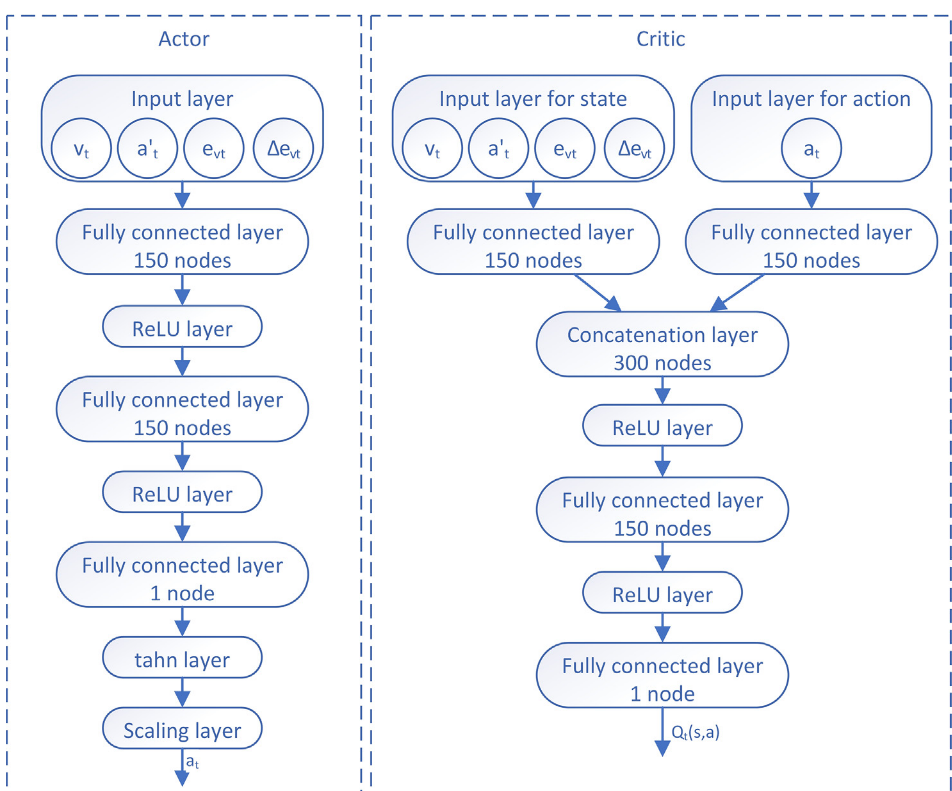

Fig. 4 Deep neural network architectures for actor and critic.

ronment. The agent maximizes positive reward at every time step if AGV movement speed approaches to the speed limit at the appropriate position of the route. This can be achieved by applying the Gaussian function with curve's peak of unit height:

$$f(x) = e^{-\frac{(x-c)^2}{2\sigma^2}} \tag{27}$$

Where c is the mean or the position of the center of the peak (in this case $c = 0$). σ – standard deviation which describes the dependency of the positive reward on the distance of the speed value from the limited ones. Such a description of the reward function component ensures that the agent will seek to maintain the maximum possible movement speed of AGV depending on the position of the driving route. Whereas the optimization of AGV movement also takes place in terms of transport time and energy consumption, the penalty is directly proportional to power generated by AGV at every time step is applied. The total reward r_t, provided at every time step t, is:

$$r_t(e_v, P) = \phi_1 e^{-\frac{e_{vt}^2}{2\sigma^2}} - \phi_2 P_t - \phi_3 \tag{28}$$

Where e_{vt} is the error between speed limits at the appropriate position of the route and actual speed at time step t, P_t is a power generated by AGV at time step t, and ϕ_1, ϕ_2, ϕ_3 are the coefficients which represent the weight of AGV movement speed, power and step cost respectively to the total reward at time step t.

3. Results and simulations

Simulation conditions of the proposed optimization method are based on real measurements of container truck routes from the quay crane to the stack at the Klaipeda container terminal. In the figure below we can see the case of movement speed measurements the container truck while transporting container on a specific path (Fig. 5) when truck mass was 17800 kg, cargo mass – 31100 kg, and total route length 822 m.

Based on the real conditions of container transportation at the terminal, a Simulink model was developed integrating deep reinforcement learning-based agent and battery-powered AGV model (Fig. 6).

The distance to the driving route (*target_distance*) is set to 822 m. By comparing this parameter with the actual distance traveled, a condition for stopping the simulation for the AGV agent is created. It also stops if the simulation time exceeds the set limit of 10000 s during the training process. Next, based on the distance traveled by AGV, the Matlab function block generates speed limits (*speed_lim*), according to which the speed error is calculated. This parameter, together with the speed and acceleration parameters obtained from the AGV model is used to generate observations i.e., four parameters vector: AGV movement speed [m/s], AGV movement acceleration [m/s2], AGV movement speed error between speed limits at an appropriate position of the route and actual speed [m/s], and derivative of AGV movement speed error between the speed limit and actual speed [m/s]. Finally, a reward calculation block is used that receives the instantaneous AGV power and speed error at the input, and using these parameters, the reward is calculated according to equation (28). To get the best training performance the values of coefficients ϕ_1, ϕ_2, ϕ_3 and σ were empirically selected 15, 0.00024, 1 and 3.5 respectively and the reward in this block is calculated by:

$$r_t(e_v, P) = 15 e^{-\frac{e_{vt}^2}{24,5}} - 0,00024 P_t - 1 \tag{29}$$

DDPG agent training was performed using Matlab R2021a and the Reinforcement learning toolbox where the training hyperparameters were selected as shown in the table below (Table 1).

The training progress of DDPG agent is shown in the picture below (Fig. 7). The training process was stopped after 200 training episodes where the average reward of 5 episodes reached 1830.72.

To test the trained agent, a simulation scenario was developed based on the real container transportation case:

1. Wait for 10 s
2. Drive 30.0 m, when the movement speed limit is 2.3 m/s
3. Stop and wait for 50 s
4. Drive 286.0 m, when the movement speed limit is 7.5 m/s
5. Stop and wait for 10 s
6. Drive 29.0 m, when the movement speed limit is 2.0 m/s
7. Stop and wait for 30 s
8. Drive 18.0 m, when the movement speed limit is 2.7 m/s
9. Stop and wait for 50 s
10. Drive 459.0 m, when the movement speed limit is 6.0 m/s

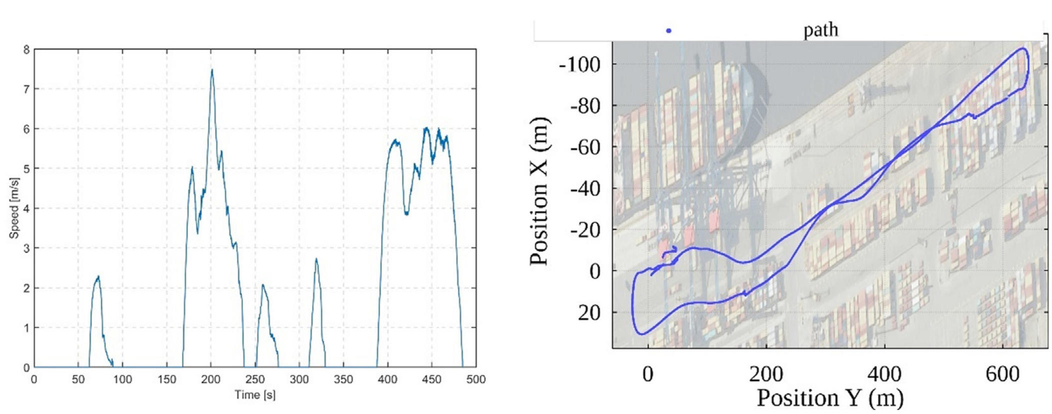

Fig. 5 Case of container transportation measurements on a container truck at Klaipeda container terminal: movement speed (left) and transportation path (right).

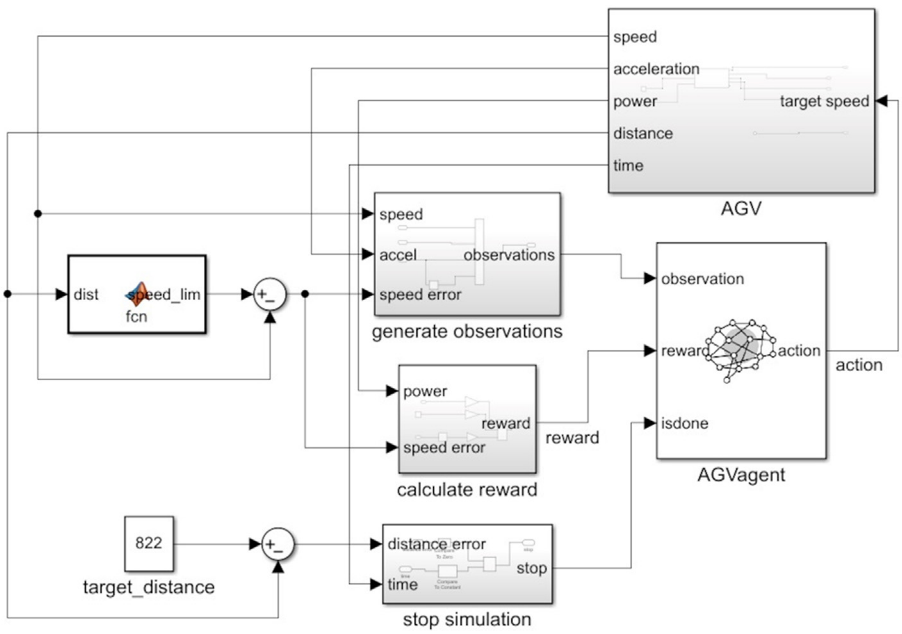

Fig. 6 Simulink model of deep reinforcement learning-based AGV control for container transportation.

Table 1 DDPG agent training hyperparameters.

Parameter	Value
Sample time	1 s
Learning rate for actor	0.001
Learning rate for critic	0.005
Batch size	128
Experience buffer length	1,000,000
Target smooth factor	0.001
Target update frequency	1
Exploration noise standard deviation	0.05
Exploration noise mean	0
Exploration noise standard deviation decay rate	0.00001

According to this scenario the total distance traveled is 822 m and waiting times are set according to the real possibilities to stop and give way to other AGVs. In the implementation of this scenario, two simulations were performed (Fig. 8):

1) without using deep reinforcement learning (DRL) controller i.e., movement speed was set to speed limit at an appropriate distance of the route
2) using DRL controller agent that automatically sets AGV movement speed according to the reward

When a DRL controller agent is used, a slower speed than the speed limit is always selected depending on the reward

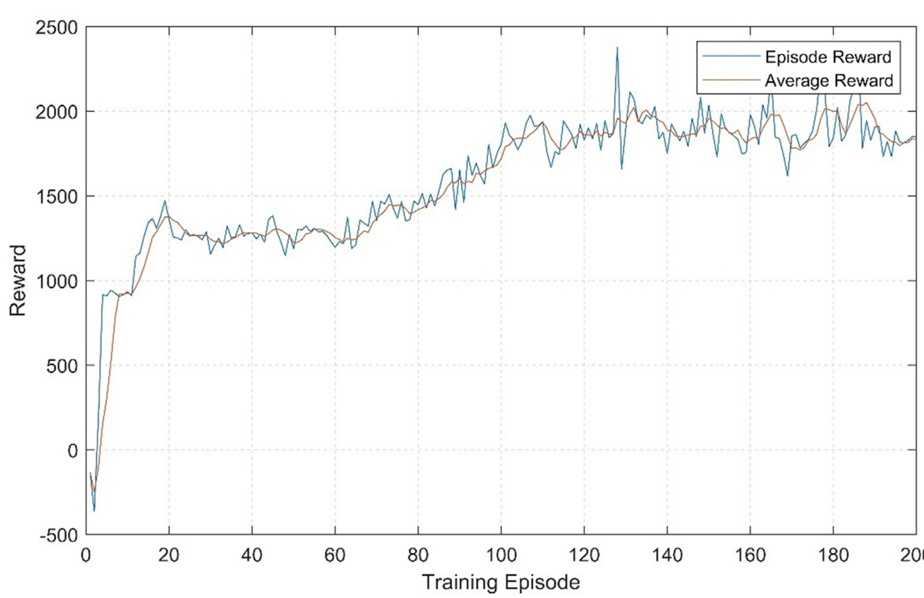

Fig. 7 Cumulative reward change during training of DDPG agent.

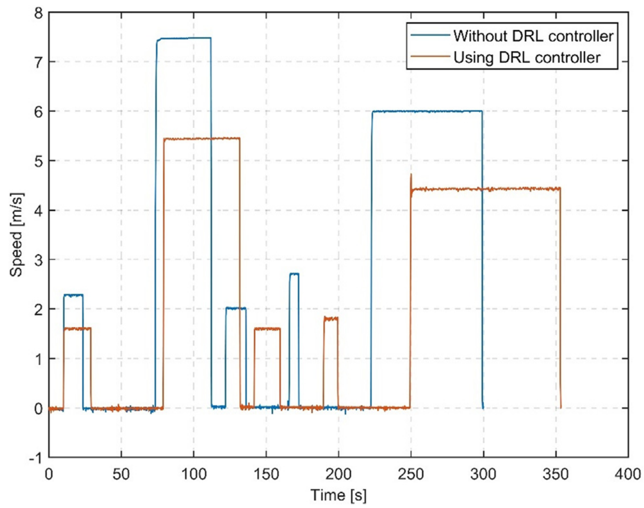

Fig. 8 AGV route simulation results without DRL controller and using DRL controller agent.

function and as a result, a travel time is automatically extended by 53 s. However, there is a noticeable reduction in energy consumption when using a DRL controller agent (Fig. 9). In this case, the when a DRL controller agent is used an energy consumption is decreased by 0.05432 kWh (i.e., from 1.18025 kWh down to 1.12593 kWh) and this is 4.6%.

During the training of the DRL agent, an optimal solution is obtained according to the constructed reward function by balancing energy consumption and time.

From the previous studies, it can be seen that most container terminal and AGV performance, travel distance optimization methods are based on scheduling, path planning and demonstrates good results in achieving a reduction in overall energy consumption of 11.65% [11], reduction in transportation distance and transportation efficiency of 4.4% and 21.8%, respectively [8], reduction in average energy cost of 25.53% [12], reduction in AGV energy consumption of 10.98% [9].

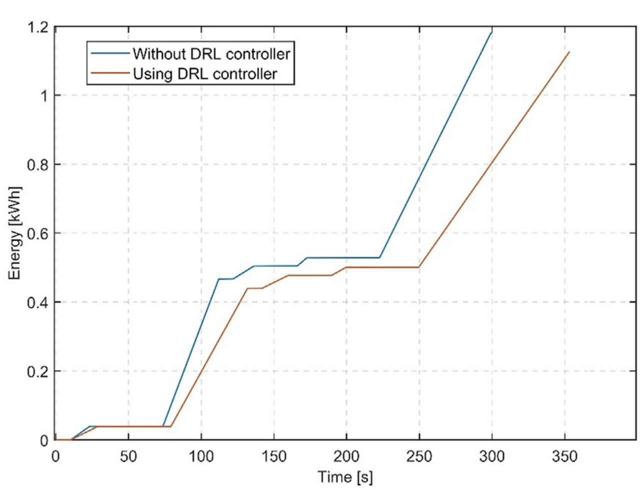

Fig. 9 AGV energy consumption comparison without DRL controller and using DRL controller agent.

In this research paper, a new DRL-based AGV energy consumption optimization method was proposed and the research results showed that by reducing energy consumption by 4.6%, it can be used as a new approach, in addition to the already mentioned scheduling and path planning, optimizing the operation of the container terminal.

4. Conclusion

Because of the increasing cargo flows in container terminals, the optimization of the energy costs and duration of container transportation becomes a key task. In this paper, we proposed an AGV simulation model and a new DRL-based AGV control method that enables energy consumption optimization. The specific conclusions are as follows:

- A simulation model of battery-powered AGV for container transportation is proposed, which allows the simulation of instantaneous energy consumption during container transportation and can be used to solve energy consumption optimization tasks in container terminals.
- A new DRL framework based DDRL agent model with integrated AGV simulation model is built. The trained DDRL agent model enabled to find energy and a time-based optimal strategy for AGV speed control during container transportation in a terminal.
- Simulation and comparison with real transportation measurements showed that the proposed DRL-based optimization approach reduces energy consumption by 4.6%. However, a transportation time increases, because of contradiction between energy efficiency and process time. Adjusting the weights of energy and time penalties used in the reward function the required weight balance of these parameters can be achieved. The proposed energy optimization method can be used as a new approach, in addition to the widely used scheduling and path planning, to optimize the operations of the container terminal.

In future works, the research problem could be expanded to find optimal strategies for autonomous vehicle movement, including context awareness and information sharing with other terminal vehicles.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would also like to thank Klaipeda University project members: prof. dr. R. Didziokas, prof. dr. E. Guseinoviene, dr. T. Eglynas, dr. S. Jakovlev, dr. M. Jusis, E. Pocevicius for deep insights and relevant contribution to the work. The authors would also grateful for the support from the Klaipeda city Containers handling company LKAB "Smelte" for providing consultations, access to the technical means, and valuable insights during the experimental phase of the project.

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the European Regional Development Fund according to the supported activity "Research Projects Implemented by World-class Researcher Groups" under Measure No. 01.2.2-LMT-K-718-01-0081, under a grant agreement with the Research Council of Lithuania (LMT).

References

[1] European Environment Agency, Greenhouse gas emissions from transport in Europe, (2022). https://www.eea.europa.eu/ims/greenhouse-gas-emissions-from-transport (accessed November 28, 2022).

[2] S. Alataş, Do environmental technologies help to reduce transport sector CO2 emissions? Evidence from the EU15 countries, Research in Transportation Economics. 91 (2022), https://doi.org/10.1016/j.retrec.2021.101047.

[3] J. Martínez-Moya, B. Vazquez-Paja, J.A. Gimenez Maldonado, Energy efficiency and CO2 emissions of port container terminal equipment: Evidence from the Port of Valencia, Energy Policy. 131 (2019) 312–319, https://doi.org/10.1016/j.enpol.2019.04.044.

[4] P.E.N.G. Yun, L.I. Xiangda, W.A.N.G. Wenyuan, L.I.U. Ke, L.I. Chuan, A simulation-based research on carbon emission mitigation strategies for green container terminals, Ocean Engineering. 163 (2018) 288–298, https://doi.org/10.1016/j.oceaneng.2018.05.054.

[5] Ç. Iris, J.S.L. Lam, A review of energy efficiency in ports: Operational strategies, technologies and energy management systems, Renewable and Sustainable Energy Reviews. 112 (2019) 170–182, https://doi.org/10.1016/j.rser.2019.04.069.

[6] Y.C. Yang, Operating strategies of CO2 reduction for a container terminal based on carbon footprint perspective, Journal of Cleaner Production 141 (2017) 472–480, https://doi.org/10.1016/j.jclepro.2016.09.132.

[7] Z. Wan, T. Zhang, M. Sha, W. Guo, Y. Jin, J. Guo, Y. Liu, Evaluation of emission reduction strategies for berthing containerships: A case study of the Shekou Container Terminal, Journal of Cleaner Production 299 (2021), https://doi.org/10.1016/j.jclepro.2021.126820.

[8] J. Chen, X. Zhang, X. Peng, D. Xu, J. Peng, Efficient routing for multi-AGV based on optimized Ant-agent, Computers & Industrial Engineering. 167 (2022), https://doi.org/10.1016/j.cie.2022.108042.

[9] L. Hu, X. Zhao, W. Liu, W. Cai, K. Xu, Z. Zhang, Energy benchmark for energy-efficient path planning of the automated guided vehicle, Science of the Total Environment. 857 (2023), https://doi.org/10.1016/j.scitotenv.2022.159613.

[10] H. Yu, Y.E. Ge, J. Chen, L. Luo, C. Tan, D. Liu, CO2 emission evaluation of yard tractors during loading at container terminals, Transportation Research Part D: Transport and Environment. 53 (2017) 17–36, https://doi.org/10.1016/j.trd.2017.03.014.

[11] C. Wang, C. Jin, Z. Li, Bilevel programming model of low energy consumption AGV scheduling problem at automated container terminal, in: 2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE), 2019, pp. 195–199, https://doi.org/10.1109/SMILE45626.2019.8965316.

[12] J. He, Y. Huang, W. Yan, Yard crane scheduling in a container terminal for the trade-off between efficiency and energy consumption, Advanced Engineering Informatics. 29 (2015) 59–75, https://doi.org/10.1016/j.aei.2014.09.003.

[13] J. Ren, C. Ye, F. Yang, Solving flow-shop scheduling problem with a reinforcement learning algorithm that generalizes the value function with neural network, Alexandria Engineering Journal. 60 (2021) 2787–2800, https://doi.org/10.1016/j.aej.2021.01.030.

[14] R.M. Abd El-Aziz, Renewable power source energy consumption by hybrid machine learning model, Alexandria Engineering Journal. 61 (2022) 9447–9455, https://doi.org/10.1016/j.aej.2022.03.019.

[15] M.H. Alabdullah, M.A. Abido, Microgrid energy management using deep Q-network reinforcement learning, Alexandria Engineering Journal. 61 (2022) 9069–9078, https://doi.org/10.1016/j.aej.2022.02.042.

[16] X. Huang, D. Zhang, X.S. Zhang, Energy management of intelligent building based on deep reinforced learning, Alexandria Engineering Journal. 60 (2021) 1509–1517, https://doi.org/10.1016/j.aej.2020.11.005.

[17] S. Malekshah, A. Rasouli, Y. Malekshah, A. Ramezani, A. Malekshah, Reliability-driven distribution power network dynamic reconfiguration in presence of distributed generation by the deep reinforcement learning method, Alexandria Engineering Journal. 61 (2022) 6541–6556, https://doi.org/10.1016/j.aej.2021.12.012.

[18] S. Touzani, A.K. Prakash, Z. Wang, S. Agarwal, M. Pritoni, M. Kiran, R. Brown, J. Granderson, Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency, Applied Energy. 304 (2021), https://doi.org/10.1016/j.apenergy.2021.117733.

[19] Z. Ren, J. Lai, Z. Wu, S. Xie, Deep neural networks-based real-time optimal navigation for an automatic guided vehicle with static and dynamic obstacles, Neurocomputing. 443 (2021) 329–344, https://doi.org/10.1016/j.neucom.2021.02.034.

[20] M.A. Dulebenets, R. Moses, E.E. Ozguven, A. Vanli, Minimizing carbon dioxide emissions due to container handling at marine container terminals via hybrid evolutionary algorithms, IEEE Access. 5 (2017) 8131–8147, https://doi.org/10.1109/ACCESS.2017.2693030.

[21] Y. Hirashima, O. Furuya, K. Takeda, M. Deng, A. Inoue, A new method for marshaling plan using a reinforcement learning considering desired layout of containers in port terminals, 2005.

[22] T. Eglynas, S. Jakovlev, V. Jankunas, R. Didziokas, J. Januteniene, D. Drungilas, M. Jusis, E. Pocevicius, M. Bogdevicius, A. Andziulis, Evaluation of the energy consumption of container diesel trucks in a container terminal: A case study at Klaipeda port, Science Progress 104 (2021), https://doi.org/10.1177/00368504211035596.

[23] R.S. Sutton, A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[24] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, (2015). http://arxiv.org/abs/1509.02971.

[25] M. Acciaro, H. Ghiara, M.I. Cusano, Energy management in seaports: A new role for port authorities, Energy Policy. 71 (2014) 4–12, https://doi.org/https://doi.org/10.1016/j.enpol.2014.04.013.

[26] K.H. Kim, S.M. Jeon, K.R. Ryu, Deadlock prevention for automated guided vehicles in automated container terminals, in: H.-O. Kim Kap Hwan and Günther (Ed.), Container Terminals and Cargo Systems: Design, Operations Management, and Logistics Control Issues, Springer Berlin Heidelberg, 2007: pp. 243–263. https://doi.org/10.1007/978-3-540-49550-5_12.

[27] G. Saldaña, J.I. San Martín, I. Zamora, F.J. Asensio, O. Oñederra, Analysis of the current electric battery models for electric vehicle simulation, Energies. 12 (2019), https://doi.org/10.3390/en12142750.

[28] A.M. Lulhe, T.N. Date, A technology review paper for drives used in electrical vehicle (EV) & hybrid electrical vehicles (HEV), in: In: 2015 International Conference on Control, Instrumentation, Communication and Computational

Technologies (ICCICCT), pp. 632–636, https://doi.org/10.1109/ICCICCT.2015.7475355.

[29] J. de Santiago, H. Bernhoff, B. Ekergård, S. Eriksson, S. Ferhatovic, R. Waters, M. Leijon, Electrical motor drivelines in commercial all-electric vehicles: a review, IEEE Transactions on Vehicular Technology. 61 (2012) 475–484, https://doi.org/10.1109/TVT.2011.2177873.

[30] NABU, Battery-electric AGVs. (2013). Https://Www.Nabu.de/Imperia/Md/Content/Nabude/Veranstaltungen/131015-Nabu-Clean-Air-Antwerp-Koetter.Pdf.

[31] P.C. Krause, O. Wasynczuk, S.D. Sudhoff, S.D. Pekarek, Analysis of electric machinery and drive systems, Wiley, 2013, https://books.google.lt/books?id = ocr9wSmSZEsC.

[32] A.N. Gent, J.D. Walter, Pneumatic tire (2006), https://ideaexchange.uakron.edu/mechanical_ideas/854.

[33] C. Zhu, X. Li, L. Song, L. Xiang, Development of a theoretically based thermal model for lithium ion battery pack, Journal of Power Sources. 223 (2013) 155–164, https://doi.org/10.1016/j.jpowsour.2012.09.035.

[34] O. Tremblay, L.-A. Dessaint, Experimental validation of a battery dynamic model for EV applications, World Electric Vehicle Journal. 3 (2009) 289–298.

[35] L.H. Saw, K. Somasundaram, Y. Ye, A.A.O. Tay, Electro-thermal analysis of lithium iron phosphate battery for electric vehicles, Journal of Power Sources. 249 (2014) 231–238, https://doi.org/https://doi.org/10.1016/j.jpowsour.2013.10.052.

[36] N. Omar, M.A. Monem, Y. Firouz, J. Salminen, J. Smekens, O. Hegazy, H. Gaulous, G. Mulder, P. van den Bossche, T. Coosemans, J. van Mierlo, Lithium iron phosphate based battery – assessment of the aging parameters and development of cycle life model, Applied Energy. 113 (2014) 1575–1585, https://doi.org/https://doi.org/10.1016/j.apenergy.2013.09.003.