https://doi.org/10.15407/knit2021.06.038
UDC 629.78

S. V. KHOROSHYLOV, Leading researcher, Dr. Sci. in Tech., Professor
ORCID.org/0000-0001-7648-4791
E-mail: skh@ukr.net
M. O. REDKA, Postgraduate
E-mail: mix5236@ukr.net

Institute of Technical Mechanics of the National Academy of Science of Ukraine and the State Space Agency of Ukraine
15, Leshko-Popelya Str., Dnipro, 49005 Ukraine

DEEP LEARNING FOR SPACE GUIDANCE, NAVIGATION, AND CONTROL

The advances in deep learning have revolutionized the field of artificial intelligence, demonstrating the ability to create autonomous systems with a high level of understanding of the environments where they operate. These advances, as well as new tasks and requirements in space exploration, have led to an increased interest in these deep learning methods among space scientists and practitioners. The goal of this review article is to analyze the latest advances in deep learning for navigation, guidance, and control problems in space. The problems of controlling the attitude and relative motion of spacecraft are considered for both traditional and new missions, such as orbital service. The results obtained using these methods for landing and hovering operations considering missions to the Moon, Mars, and asteroids are also analyzed. Both supervised and reinforcement learning is used to solve such problems based on various architectures of artificial neural networks, including convolutional and recurrent ones. The possibility of using deep learning together with methods of control theory is analyzed to solve the considered problems more efficiently. The difficulties that limit the application of the reviewed methods for space applications are highlighted. The necessary research directions for solving these problems are indicated.

Keywords: *spacecraft, deep learning, navigation, guidance, control, artificial neural network, reinforcement learning, landing, hovering.*

INTRODUCTION

At this time, artificial intelligence methods attract great interest of researchers and practitioners all over the world [2], which is largely due to the impressive results obtained using deep learning (DL) techniques [7]. DL has rapidly evolved and showed promising results in solving complex tasks, finding non-trivial solutions of existing problems [47]. DL-based systems are already successfully used in practice in various fields, for example, in computer vision [50], natural language processing [56], autonomous driving [42], robotics [38], etc. Meanwhile, space control systems have been designed mostly based on classical methods, for example [1, 16]. However, the developers face the problems of adaptability, robustness, and autonomy when they attack new problems of space exploration using conventional techniques.

Complex tasks of an orbital service, such as releasing a payload or capturing non-interacting targets, are accompanied by rapid changes in attitude and mass parameters of the SC, which can lead to unstable motion and tumbling of the satellite [51]. In such conditions, the driving modes and mass characteristics are unpredictable. In these cases, con-

Цитування: Khoroshylov S. V., Redka M. O. Deep learning for space guidance, navigation, and control. *Space Science and Technology.* 2021. **27**, № 6 (133). С. 38—52. https://doi.org/10.15407/knit2021.06.038

ventional attitude control techniques cannot always stabilize the satellite since they depend on the mass parameters of the plant.

The desire to expand the human presence in the near moon space determines the demand for "light" automated onboard processes of SC [33]. Finding computationally efficient guidance strategies is challenging for nonlinear dynamic objects. Many conventional approaches rely either on simplifying assumptions in the dynamic model or significant computational resources.

The necessity to operate successfully under uncertainty is becoming evident for future space missions that are being developed to explore various bodies in deep space. High-precision SC attitude control, which is robust to emerging perturbations and uncertain dynamics, is very important to continue successful space flights into deep space.

Future full-scale missions to Mars will require advanced guidance algorithms that can adapt to more stringent requirements, such as autonomous landings in selected regions with maximum precision and cost-effective flight trajectories [18]. Missions to the Moon and Mars require a perfect navigation system and control algorithms for the descent phase. Such algorithms should estimate the SC state vector using input data from the array of sensors and generate the required control commands to ensure an accurate landing in an optimal way. This problem was traditionally solved offline when optimal landing trajectories were calculated in advance on the Earth and then used in onboard control algorithms, which violates the autonomy of the mission. Autonomous fault detection and recovery capability are also essential for future manned missions to Mars.

The currently used methods for maneuvering near asteroids require extremely accurate knowledge of the environment model and accurate knowledge of the SC position before the maneuver [16]. Acquiring this knowledge is both time-consuming and expensive. This leads to a delay of several months between the arrival of the SC to the asteroid and the moment when it is possible to start safely performing maneuvers in its proximity. Autonomous operations in close proximity (including hovering and landing) in a low-gravity asteroid environment are particularly challenging.

Mapping and navigating around a small unknown body continue to be an extremely interesting and exciting challenging task in space exploration [5]. Traditionally, the trajectory of a SC for mapping missions is calculated by a human expert, which requires hundreds of hours of human time to control the navigation process and orbit selection. While the current methodology has yielded satisfactory results for previous missions (e.g., Rosetta, Hayabusa, and Deep Space), current requirements for mapping missions are expanding, requiring additional autonomy during the mapping and navigation process for the SC.

The high cost of space missions has prompted several space agencies to intensify the development of autonomous SC control systems [21]. Learning agents represent one of the possible ways in which an autonomous SC can adapt to changing equipment capabilities, environmental parameters, or mission objectives while minimizing dependence on interventions from the Earth.

New requirements and tasks in the field of space exploration, as well as significant advances in applying DL technologies in other areas, have inspired research to apply these methods for space applications, and many interesting results in this field have been published in the past several years.

The goal of this survey is to analyze the results of recent work on the adaptation of DL methods for space navigation, guidance, and control tasks. To the best of our knowledge, this is the first specialized review devoted directly to this specific field with this scope. However, there are reviews close to this topic [24, 34, 59], which can be useful for deepening or expanding knowledge in this field.

The remainder of the article is organized as follows. Section 1 introduces the basic concepts of DL that are necessary to understand the material of the next sections for readers who are not familiar with this subject. Section 2 reviews publications related to the topic of guidance and navigation. Sections 3 and 4 analyze publications on SC attitude and relative control, respectively. The problems of landing on space bodies and flights in their proximity are considered in Section 5. Mission planning and high-level decision-making tasks are reviewed in Section 6. The main problems of DL implementation and possible ways of their solution are summarized in Conclusion.

1. DEEP LEARNING

Machine learning is a subset of artificial intelligence methods used to develop algorithms that can solve a problem by searching for patterns in a variety of input data [35].

There are two main paradigms of machine learning:

1. Supervised learning is the task of learning a function that maps an input to an output based on example input-output pairs. The cost or loss function, as a rule, is determined through the average error of the algorithm responses for all examples. Supervised learning includes classification and regression problems.

2. Unsupervised learning is a type of algorithm that learns patterns from untagged data. This includes the tasks of clustering, anomaly detection, latent variable models, etc.

There are also so-called semi-supervised learning methods that take the place somewhere in between supervised and unsupervised learning. Reinforcement learning (RL) [41] is an example of semi-supervised learning. According to the RL setup, an intelligent agent learns by interacting with some environment. The separation line between the environment and the agent is rather vague, but from the perspective of the tasks analyzed in this article, the agent can be considered as a control system, and the environment can be considered as a plant.

The environment is usually modeled in the form of a Markov decision process (MDP). In this regard, many RL algorithms use dynamic programming methods [20]. The main difference between the classical methods of dynamic programming and RL algorithms is that the latter do not imply knowledge of the accurate mathematical model of the MDP and are especially expedient when it is difficult to apply exact methods. The goal of the RL is to find an optimal sequence of actions of the agent, called a policy, which maximizes the reward function. Similarly to the control law in classical control theory, the policy often maps the state vector and control actions.

Machine learning methods based on artificial neural networks (ANN) [7] are called DL. New ANN architectures have largely determined the latest advances in DL. For example, convolutional neural networks (CNN) are effective for the analysis of visual data [32]. CNNs are based on convolution kernels or general-weighted filters that slide over the input and provide equivalent responses, known as feature maps. Recurrent neural networks (RNN) are used to process input data sequences of variable length [19]. The connections between RNN nodes form a directed graph in a time sequence. This allows them to demonstrate dynamic behavior over time using their internal state (memory).

Different approaches are used to improve the effectiveness of DL, for example, transfer-learning (TL) [53] and meta-learning (ML) [49]. TL allows the system to improve the learning of a new task by using the knowledge gained earlier in solving a similar task. ML is based on the use of such metadata, which makes automatic learning more flexible and improves the efficiency of existing algorithms.

2. NAVIGATION AND GUIDANCE

Reference [31] presents a new method for attitude determination using color images of the Earth obtained with a visible range camera. A conventional earth camera is used to coarsely determine orientation by detecting the edge of the earth, and therefore only provides a rough 2-axial orientation. In contrast, the proposed method can provide information to determine the orientation about three axes by recognizing the earth pattern with an accuracy of fractions of a degree and then comparing the detected earth pattern with the global map. In addition, this method can be implemented on the basis of a detector system consisting of an inexpensive optical color camera and a single onboard computer. This reduces the size, weight, and cost of the system.

To demonstrate this method in space, a sensor system called the “Deep Learning Attitude Sensor” has been developed [31]. The resulting images are quickly analyzed using state-of-the-art real-time image recognition algorithms. The algorithm demonstrates good performance in various images. Image processing time to determine the orientation is less than 6 seconds. It is noted that the ANN architecture should be selected depending on the performance of on-board computers. In the future, it is planned to apply one of the U-net image segmentation methods, which is a CNN for fast and accurate image segmentation.

A new approach called deep guidance is investigated in Ref. [22]. The authors use deep RL to learn guidance policies instead of handcrafting them. The deep guidance technique includes a trained guidance policy that generates speed commands to the conventional tracking controller. The application of the deep RL in conjunction with control theory methods reduces the training load and facilitates the transfer of the trained system from simulation to reality. Simulations and experimental studies of scenarios for position tracking and docking of the SC were carried out to test the feasibility of the proposed approach. The results show that such a system can be fully simulated and transferred into real-world conditions with an acceptable loss of performance without any additional tuning. Apparently, this is the first experimental demonstration of an artificial intelligence application to control the motion of a SC.

Reference [18] proposes a new adaptive guidance system developed using meta-RL. In this work, the principles of meta-RL are used to obtain two laws of adaptive guidance. The first one is designed to control the SC during the descent to Mars, and the second one is suitable for landing on small bodies, such as asteroids. Guidance laws take the form of a global policy in the state space, determined by the deployment region and the places of possible landing. This global policy maps the estimates of the lander state vector in the target coordinate system and the thrust control vector. The system architecture includes the RNNs for the implementation of the control law and the cost function approximator. The RNN allows the obtained algorithms to adapt in real time to environmental disturbances acting on the SC.

Policy optimization involves modeling the interactions between the agent implementing the policy and the environment over many episodes with randomly generated initial conditions that cover possible scenarios of operation [18]. It is important to note that environmental parameters such as state error, lander wet mass, and disturbance forces vary between episodes. The optimized strategy adapts to these parameters in real time. The effectiveness of the policies of recurrent and non-recurrent RL agents is analyzed in comparison with conventional guidance algorithms in four complex tasks with unknown and very volatile dynamics. Such tasks include a safe landing on Mars with an accidental engine failure and landing on an asteroid with an unknown model of the environment.

Using a series of experiments, it has been demonstrated that the meta-RL guidance outperforms the conventional feedback control algorithm with optimal power consumption [18]. In addition, it is shown that in a subset of experiments, the adaptive meta-RL guidance outperforms the non-adaptive guidance law optimized using standard RL. It should be noted that the RL policy is optimized in the same environment as the meta-RL policy. The possibilities of an optimized ML policy for obtaining and implementing the guidance law using observations consisting only of Doppler radar altimeter measurements during landing on Mars and LIDAR altimeter measurements during landing on an asteroid are demonstrated. This example illustrates the capabilities of the simultaneous solution of guidance and navigation tasks.

Reference [15] presents a new guidance law, which uses only measurements of the seeker viewing angles and their rates. The policy is optimized using meta-RL, and its effectiveness is demonstrated by simulating the final phase of exoatmospheric interception. It is important to note that guidance law does not require range estimates, making it particularly suitable for passive seekers. Optimized guidance law maps the stabilized view angles and velocities of the finder and the thrust of the control thrusters of the missile. Meta-RL allows the optimized policy to adapt to the acceleration of the target. It is demonstrated that the RL policy provides better guidance performance compared to the algorithms that use accurate target acceleration knowledge.

Each training episode is based on a scenario of interaction with random parameters [15]. A scenario of high-altitude interception of a maneuverable ballistic missile is considered, in which the intercept missile must kinetically destroy the target by a direct hit (miss less than 50 sm). The interaction scenario has been significantly simplified. First, only the final phase of the interception is modeled. Secondly, realistic ballistic trajectories of the missile and target are not generated, and gravity is also neglected. Note, however, that it is common practice to neglect gravity when initially developing a new guidance law.

Third, target separation and angular motion control are not considered. In this case, the interaction scenario assumes that the target has an initial speed advantage, namely the initial missile and target speeds are 3000 m/s and 4000 m/s, respectively. The target performs a random bang-bang maneuver during interception. This is a realistic maneuver for a descent vehicle in order to avoid interception in a way that does not radically change the trajectory of the vehicles entering the atmosphere. The ratio of the thrust-to-target ratio of the missile to the target is 2:1. The RL policy provides improved guidance accuracy and fuel efficiency.

The effectiveness of the RL policy [15] is achieved by using RNNs to approximate the policy and the cost function, which enables the policy to adapt to a specific target maneuver in real time. In particular, the hidden states of the recurrent layers change in different ways in response to target maneuvers for each specific interaction, which makes it possible to form control actions taking into account the specifics of the maneuver. In contrast to the zero-effort miss policy [57], where the state filter evaluates accelerations, the RL policy adapts to the target maneuver in real time. The optimized policy is computationally efficient, requires minimal memory size, and is compatible with modern onboard processors.

In Ref. [17], adaptive integrated guidance, navigation, and control system was developed for maneuvering in the proximity of asteroids with unknown environmental dynamics, with initial conditions covering large launch areas, and without knowing the model of the asteroid shape. The system is implemented as a policy optimized using meta-RL. The SC is equipped with an optical finder which is fixed either on a certain feature of the surface, or on the reflected light from a targeting laser, or an active beacon. The policy directly maps observations, including the finder angles and LIDAR distance, and the thrust commands. The policy is implemented in the form of an RNN, which can adapt in real time to both external disturbances acting on the agent and internal disturbances such as actuator failures and variations of the center of mass.

The guidance system was validated by modeling landing maneuvers using a simulator with six degrees of freedom [17]. The simulator randomizes asteroid characteristics such as solar radiation pressure, density, rotation speed, and nutation angle, requiring appropriate adaptation of guidance and control algorithms. The authors demonstrate the robustness of the system to failure of the actuators, displacement of sensors, as well as variation of the inertia and center of mass of the SC.

The concept of maneuvers for performing operations in the proximity of asteroids is proposed in Ref. [17], which is compatible with the developed guidance system. In future studies, the authors plan to focus on search technologies that rely on terrain features rather than beacons and replace the LIDAR rangefinder with binocular vision. Another direction of future work may study the possibility of positioning the SC on the collision triangle with the target, which can increase both fuel efficiency and robustness with respect to fast rotations of asteroids. The approach discussed in this paper can also be applied to orbital rendezvous and landing on the Moon, especially for the Moon missions, where the landing site is already indicated by beacons.

Reference [33] is motivated by the demand for “light” automated onboard processes of SC to expand the human presence in the near Moon space. The controller proposed in this work uses nonlinear equations of motion, but this does not lead to a large additional load on the on-board computer. However, such guidance principles can leverage high-performance computations by decoupling the learning process from the resulting controller. Practical examples demonstrate the flexibility of the RL approach and the possibility to use it for tasks requiring higher guidance accuracy. The controller provides computationally efficient onboard guidance based on multiple bodies. The ANN controller demonstrates robustness to the reference geometry variations and can generalize past experience for solving new tasks. In addition, the proposed approach separates the learning agent from environmental dynamics, which provides model-free guidance.

Reference [11] introduces an approach for performing precise landing on the planets of the solar system, based on the methodology for the trajectory formation using potential functions. The theory of extreme machine learning is used to develop a single-layer feed-forward ANN, which learns to map the current position of the SC and the optimal velocity field necessary to form a path to the planet's surface, taking into account fuel economy. Extreme learning methods provide fast and accurate learning as well as good generalization. The network is trained in an open loop using economical trajectories that are generated numerically using pseudospectral methods. Once tested and verified, the ANN becomes a critical element in the loop of the linear guidance algorithm. In particular, a linear-quadratic controller (LQR) is used to track the field of optimal speed, which is determined to be attractive for a landing target. Monte Carlo simulations show that the algorithm provides a low residual pointing error of less than one meter in position and less than 0.9 m/s in velocity.

The adaptive RL-based guidance algorithm for real-time trajectory tracking is designed in Ref. [14] for reliable, cost-effective, and accurate landing on Mars without needing it to build on Earth first. The results of the Monte Carlo simulations show that the algorithm is capable of autonomously providing movements along close to optimal trajectories with minimal fuel consumption and with an accuracy exceeding the accuracy of past and future missions to Mars. The RL-based guidance algorithm demonstrates a high degree of flexibility and can easily adapt autonomous retargeting while maintaining accuracy and fuel efficiency. Although RL and other similar machine learning methods have previously been applied to aerospace guidance and control problems, this work is the first attempt to apply RL to the problem of autonomous planetary landing.

The adaptive guidance algorithm [14] based on RL allows the SC to be trained while performing the best landing by selecting the sequence of acceleration / thrust commands that provide economical trajectories and the necessary accuracy in terms of the desired final position and speed (soft landing). This problem was solved by developing an ANN guidance algorithm representing the landing problem in the form of an MDP. The training procedure tunes the ANN weights that maximize the expected performance criterion. The latter ensures that only those control actions are selected that move the descent vehicle to the desired location with maximum accuracy and with minimum fuel consumption. It is important to note that the proposed algorithm does not require any reference trajectory. As a result of the learning process, the network determines autonomously the optimal landing algorithm using the current position and speed information provided by the navigation system. Moreover, the system can learn the optimal landing in the presence of adverse factors such as environmental disturbances, noise, and delays in sensors and actuators.

3. ATTITUDE CONTROL

Reference [9] presents a framework for developing an adaptive SC attitude controller using deep RL. It is shown that the controller efficiently performs large-angle slew maneuvers with industry-standard pointing accuracies. The controller can adapt to various disturbances that were not presented at the training stage and does not depend on the parameters of the SC, even if it was trained on a different configuration with different parameters. Different RL methods and reward functions are investigated to improve the control accuracy. It has been demonstrated that these controllers can be trained on a modern personal computer.

A state-of-the-art single-actor RL algorithm is implemented and applied in a designed simulation environment [9], where a trained agent achieved the industry-standard accuracy in a relatively short training time. The robustness of the agent to uncertainties of the environmental conditions was tested in four different test scenarios, which are designed to simulate different conditions that the agent may encounter in space. The agent successfully adapted to all perturbation tests performed, demonstrating results close to time optimal. The ability of the agent to be robust to conditions that were not explicitly used during training makes it possible to substantiate the possibility of using RL-based controllers on real SC. The results also suggest that it is possible to use one "basic" control algorithm for a wide range of satellites, which allows increasing the constellation of autonomous SC. This is a necessary step for space exploration of the future. The results of this work can be expanded by applying the latest achievements in distributed RL in order to use data generated by a constellation of satellites to solve SC attitude control problems.

Reference [51] investigates rapid changes of orientation and mass parameters that SC encounters performing complex tasks, such as dropping-off a payload or capturing an object. This work proposes a new algorithm for attitude control based on the deep RL. A three-dimensional modeling environment has been developed that simulates the SC attitude variation in real time, taking into account the control torques. An ANN model based on a segmented weighted reward function is proposed. The ANN takes the parameters of the SC orientation as an input and outputs a discretized control torque.

The deep Q-learning algorithm [52] was used to train the agent for the task SC attitude control. Simulation experiments show that thanks to continuous self-learning and self-improvement, the deep RL agent gradually learns to restore the SC attitude after unknown disturbances. The proposed algorithm is compared with the proportional-derivative (PD) controller and the backstepping controller. The PD controller cannot restore the SC attitude due to its dependence on inertial parameters. The backstepping controller is robust against mass uncertainty but can only handle a constant control cycle. Compared to these two conventional controllers, the deep RL algorithm provides competitive performance in the presence of mass uncertainties and allows the control loop to be varied during the learning phase. The proposed mechanism makes it possible to implement intelligent control and can serve as a technical basis for SC orbital service.

Reference [8] presents a framework for designing a discrete neural SC attitude controller using RL without high-performance computations. Quasi-time-optimal constrained control algorithms are obtained, capable of providing attitude accuracy significantly exceeding industry standards. The control tests of the agent performing SC large-angle slews in the developed modeling environment are also presented. The selected reward function allows the agent to improve the accuracy of attitude control beyond the minimum specified requirements. This feature illustrates the advantages of RL over classic control methods. The ability of the controller to understand long-term dependencies in processes in the presence of external disturbances or other constraints makes it possible to improve control efficiency and performance. In the short term, it is of interest to consider the capabilities of distributed RL. For example, distributed RL can be used to train agents online using data from a constellation or swarm of satellites.

Reference [4] deals with the task of SC optimal attitude control using a minimum number of thrusters. Three possible solutions to this problem are presented:

1) an easy-to-implement logic-based controller;

2) a projective controller trying to approximate ideal continuous control as accurately as possible;

3) an optimal neural network predictive controller (NPC) that minimizes the total impulse during the maneuver.

The NPC includes an RNN to predict the state vector in the finite time horizon of the optimization. Due to the fact that the considered system has discrete inputs, the backpropagation algorithm traditionally used for continuous systems is not applicable for the case considered in this article. In this paper, the NPC is adapted for binary input systems using a robust genetic algorithm to optimize the receding horizon. An automatic selection of the parameters of the cost function is proposed, which improves the performance of the NPC and reduces the number of adjustable parameters to one. In addition, the multilayer perceptron is trained offline using data obtained under optimal control. This approach allows designers to replace the cost of a function-based algorithm that requires intensive CPU computations with a much less computationally expensive meta-model.

The performance of the NPC is compared with the proposed logic and projective control algorithms for 12U CubeSat [4]. The NPC is the most effective from the point of view of the total impulse, the least sensitive to the choice of parameters, and has the same settling time. Multilayer perceptron control drastically reduces the computing resources required online, with control performance comparable to the NPC. A comparative analysis of the considered controllers showed that the NPC allows the system to save up to 25 % of fuel for the de-tumbling task and up to 36 % of fuel for slew maneuvering.

The ability of an RL agent to find the optimal control strategy for SC attitude control is studied in Ref. [48]. Two main types of attitude control systems are considered. First, the general problem of attitude control is investigated for the case of a full set of reactive actuators with restrictions on their control torques. Then, reaction wheels were used for attitude control with additional constraints. To obtain the attitude control policy, the proximal policy optimization algorithm (PPO) [44] was used to train the RL agent. To ensure robustness, the satellite inertia matrix is considered unknown to the agent and is randomly selected for each new episode of simulation. Since the plant is non-linear, curriculum learning is used to increase training efficiency. The RL-based controller is compared to the well-proven control strategy, known as the quaternion rate feedback (QRF) controller.

The nominal performance and robustness to uncertainties in the dynamics of the system are investigated [48]. The RL-based agent adapts to any SC mass without re-training it. In the mass range of 0.1 to 100,000 kg, the RL agent provides 2 % better control performance than the QRF controller tuned for the same mass range, and its performance is similar to the QRF controller tuned specifically for a given mass. In the case of the reaction wheels, the trained RL agent provides 25 % higher reward function values than the tuned QRF controller.

Reference [55] proposes an approach based on deep RL to increase the adaptability and autonomy of the satellite control system. It is a model-based algorithm that can find solutions in fewer training episodes than model-free algorithms. The simulation shows that when the classical control fails, this approach can find a solution and achieve the goal in one hundred training episodes. To optimize the policy, heuristic search is used to avoid local optima inherent in gradient methods. Compared to classical control methods, this approach does not require prior knowledge of the parameters of the satellite and its orbit but can be adapted to different situations based on the data obtained. To improve the efficiency of adaptation to various types of satellites and various tasks, it is proposed to use transfer learning.

Reference [58] is devoted to the model-free attitude control of a rigid SC in the presence of saturation of the control torque and the action of external disturbances. A model-free deep RL controller is proposed, which can continuously learn using feedback signals from the plant and implement high-precision SC attitude control without re-adjusting the controller parameters. Taking into account the continuity of the state and control action space, the twin delayed deep deterministic policy gradient algorithm (TD3) [10] is applied using “actor-critic” architecture. TD3 is more efficient than the Deep Deterministic Policy Gradient (DDPG) algorithm.

Nevertheless, the learning process is time-consuming because the TD3 agent optimizes the policy by interacting with the environment without using any prior knowledge [58]. To mitigate this problem, the PID-Guide TD3 algorithm is proposed to speed up learning and improve the convergence of the TD3 algorithm. Given that RL is difficult to implement in real conditions, a method of preliminary preparation for deployment and fine tuning is proposed. The method allows the agent not only to save training time and computational resources but also to quickly achieve good results. The experimental results show that the RL controller can implement high-precision attitude stabilization, as well as the required trajectory tracking with a high response speed and small overshoot. The proposed PID-Guide TD3 algorithm has a faster learning rate and is more robust than the TD3 algorithm.

Reference [36] investigates the attitude motion of a SC capturing non-cooperative targets. RL is used to stabilize the SC attitude under conditions of rapid variation of attitude and mass parameters. An ANN model has been built to output a discrete control torque for the SC control. An environment for modeling the SC dynamics has been developed, and the ANN is trained in this environment using the deep Q-learning algorithm. The agent receives a reward if the satellite is successfully stabilized. Simulation shows that when the learning process is repeated, the ANN gradually learns to restore the SC orientation after an unknown disturbance. On the contrary, the traditional PD controller did not cope with this task due to its dependence on mass parameters. This method of SC attitude control demonstrates significant versatility and has great potential for intelligent control of SC performing complex tasks in the future.

The goal of Refs. [28—30, 40] is to develop an effective algorithm for SC intelligent control based on RL methods. To increase the RL efficiency, a statistical model of SC dynamics based on the concept of Gaussian processes is used. On the one hand, such a model allows authors to use a priori information about the plant, and it is sufficiently flexible, and on the other hand, it characterizes uncertainty of the dynamics in the form of confidence intervals, which can be clarified during the SC operation. In this case, the task of studying the state — control action space is to obtain such measurements that reduce the boundaries of confidence intervals. As a reinforcement signal, a well-known quadratic criterion is used to take into account both accuracy requirements and control costs. The control actions are found based on the RL using the algorithm of the policy iterations. To implement the controller and evaluate the cost function, ANN approximators are used.

Guarantees of stability of the SC motion taking into account the uncertainty of the dynamic model, are obtained using the method of Lyapunov functions [30]. The cost function is chosen as a candidate for the Lyapunov function. In order to simplify stability verification on the basis of this methodology, the assumption about Lipschitz continuity of the dynamics of the plant was used, which made it possible to use the Lagrange multiplier method to find control actions taking into account the constraints formulated using the upper bound of uncertainty and Lipschitz dynamics constants. The efficiency of the proposed algorithm is illustrated by the results of computer simulations. The approach makes it possible to develop control systems that can improve their performance as data is accumulated during the operation of a specific object, and allows developers to reduce the requirements for its elements (sensors, actuators), not to use special test equipment, and reduce time and cost of the development.

4. RELATIVE CONTROL

The policy for docking maneuvers with six degrees of freedom was developed on an RL basis and implemented in the form of a feedback control law in Ref. [6]. RL provides a feasible approach for reliable, autonomous maneuvers under uncertain conditions with low computational costs. An RL algorithm is used to obtain a docking policy in a certain region of the state space of the plant with six degrees of freedom, trying to minimize the performance criterion and control costs. The simulation results of rendezvous and docking maneuvers for the Apollo mission demonstrate that the capabilities of the resulting policy are comparable to the results obtained by conventional optimal control methods. As for directions for future work, specific problems and their possible solutions, as well as the advantages and disadvantages of docking algorithms based on RL, are discussed. This work can serve as a basis for further investigation of the RL-based control for rendezvous operations under uncertain conditions.

Reference [46] synthesizes an adaptive neurocontroller for the formation flying of two SCs in low near-Earth orbit. One of the SCs is considered to be controllable, the second one is uncontrollable with an unknown ballistic coefficient. The controlled SC is capable of changing its cross-section within certain limits, as well as making impulse maneuvers. The main approaches to solving this problem are discussed. Two ANNs are introduced, and their optimal structure is found. The task is to adjust two parameters: the ballistic coefficient of the uncontrolled SC and the density of the atmosphere that are input to the control ANNs but unknown to the controlled SC. This problem is solved by a non-gradient optimization method.

Reference [27] approximates optimal relative control of an underactuated SC using RL and studies the influence of various factors on the performance of such a solution. The problem of in-plane SC relative control using only control actions applied only in-track direction is considered. This approach makes it possible to reduce the propellant consumption of the thrusters and to simplify the architecture of the control system. However, in some cases, methods of the classical control theory do not allow obtaining acceptable results. In this regard, the possibility of solving this problem by the RL methods has been investigated. This approach allows designers to find control algorithms, which are close to optimal, as a result of interactions of the control system with the plant, using a reinforcement signal characterizing the quality of control actions.

The RL-based search for control actions is made using the policy iteration algorithm [27]. This algorithm is implemented using the “actor-critic” architecture. Various options for the “actor-critic” representation using ANN approximators are considered to implement the control law and obtain the value function estimates. It is shown that the accuracy of the optimal control approximation depends on a number of features, namely, the successful structure of the approximators, the method for updating the parameters of the ANNs, and the parameters of the learning algorithm. The approach makes it possible to solve the considered class of control problems for controllers of different structures. Moreover, the approach allows the control system to improve its control algorithms during the SC operation.

5. LANDING AND HOVERING CONTROL

An adaptive landing algorithm is presented in Ref. [45], which learns to form the optimal thrust commands to ensure an accurate landing on the Moon using images and altimeter measurements as input data and the obtained experience. A new approach based on meta-RL is proposed, which combines intelligent guidance and navigation functions, providing a complete solution to the problem of landing on the Moon based on the obtained images. In particular, a simulation environment has been developed that combines the dynamics of the system and images obtained from the on-board cameras. This is achieved by merging a Python simulator with a ray tracer (such as Blender) that generates accurate images using lunar digital terrain models and a physics rendering engine. The images are then used to update the policy in real time using RL. The advantages of the latest achievements in the field of CNN and RNN for image processing and RL for policies are used to develop an agent for performing an optimal soft landing.

Considering the failures of the actuators and the uncertainty of the atmospheric parameters, a new active fault-tolerant algorithm for controlling the descent to Mars using an ANN and adaptive inversion of the model is presented in Ref. [23]. The ANN is used to detect failures and isolate them online. Then, an adaptive ANN PID controller based on the inversion of the structural adaptive model was developed for fault-tolerant control of descent to Mars. When a malfunction is detected in the actuator, the system automatically activates the ANN PID controller replacing the traditional PID controller. The error between the output of the reference model and the output of the attitude control system is corrected in such a way as to provide the required dynamic properties of the descent vehicle. The stability of the closed-loop of the control system is investigated using the Lyapunov functions. The effectiveness of the developed algorithm is illustrated by the results of computer simulation. Considering that the detection and isolation of failures increase the computational load on the control system, in future works, it is advisable to consider the possibility of fault-tolerant control without the need to explicitly perform such operations.

In Ref. [13], a new nonlinear controller for hovering operation under low gravity conditions of the asteroid environment was developed using RL. The controller is robust enough for accurately hovering in unknown environments. The controller capabilities are limited only by the maximum thrust requirements of the environmental conditions. The robustness of the controller is demonstrated by simulating precise hovering in multiple environments that were unknown during the policy optimization. The environment is modeled using non-uniform rotation and non-uniform gravity field. Models of the shape of the asteroid Itokawa were used for modeling. The performance of the RL control is compared with the PD and LQR controllers. An approach based on optical finders is presented to estimate the SC state vector relative to a landmark on the asteroid surface. The current state of the SC is accurately estimated using only a camera and laser rangefinder.

The policy with six degrees of freedom to control hovering over an asteroid was optimized in Ref. [16] using meta-RL. The ANNs of the policy and cost function include recurrent hidden layers, and the policy network additionally has an input module consisting of convolutional layers. The policy maps the pulsed LIDAR measurements and commands of the thrusters. This policy allows the SC to hover in a fixed position and with a given orientation relative to the reference frame fixed with the asteroid. It is important to note that the policy does not require position and velocity estimates and can also operate in environments with unknown dynamics and without an asteroid shape model and navigation aids. During the optimization, the agent encounters a new, randomly generated asteroid for each episode, ensuring that it is not familiar with the shape and texture of the asteroid, as well as with the environmental dynamics. The experiments demonstrate that the policy can be used for the asteroid with new characteristics. The hover controller simplifies mission planning since the SC can immediately perform hovering right after arriving at the asteroid. This, in turn, simplifies the creation of the shape model and allows remote sensing mapping of resources immediately upon arrival at the target asteroid.

Reference [5] presents a framework for optimizing the tasks of autonomous visualization and mapping as a partially observable MDP. A new environment for simulating orbital small body mapping is developed. It is demonstrated that policies trained with this MDP formulation are able to maximize the map quality while autonomously selecting orbits and controlling imaging tasks. The integration of deep RL modules into the classical SC software systems and some problems that can be encountered in this case are discussed.

The authors of Ref. [54] used deep RL to control the SC around a small celestial body, the gravitational field of which is unknown. It is assumed that the small body is a 3D ellipsoid, and its density and dimensions are uncertain within a wide range. Experiments were carried out with different systems of perception of the SC, highlighting light neuromorphic systems for detecting optical flow. It is demonstrated that even in such a highly uncertain environment and with limited sensory capabilities, the proposed approach can provide a control strategy that allows the SC to hover over the asteroid surface with little residual drift. The SC orbiting in an unknown gravitational field due to the complex rotation of the body modeled as MDP. A direct policy search algorithm was used to find control capable of keeping the SC hovering at a given point. In contrast to previous studies, the 3D ellipsoid and a number of different sensor inputs are considered. The proposed approach allowed the authors to find policies that can also minimize drift when elementary motion sensors are the only proprioceptive sensors on-board. This result is the first step towards obtaining visual-aid-based low-gravity landing algorithms.

6. MISSION PLANNING AND DECISION-MAKING

The possibility of using deep architecture to control all or part of the SC on-board decision-making system in navigation and control tasks is studied in more detail in Ref. [43]. Deep ANNs are used to form optimal control actions during landing at a given point and obtain accurate information about the state of the plant. The trained deep ANN demonstrates close to optimal landing results. These results make it possible to develop an on-board real-time optimal control system capable of generating optimal actions for large sets of possible initial states. The article shows how deep ANNs can be trained to implement optimal state feedback control for a number of continuous deterministic nonlinear systems that are of interest to the aerospace industry. The capabilities of trained networks are not limited to predicting the optimal state feedback in a subset of the state space used during training but are also able to generalize these results to cases that go far beyond the training data. This feature allows authors to assume that the ANN has learned the basic model that is the solution of the Hamilton — Jacobi — Bellman equation. The depth of the ANN strongly affects the obtained results. It is noteworthy that small networks trying to approach the optimal state feedback cannot satisfactorily approximate its complex structure. Errors caused by the use of the trained ANN do not significantly affect the final value of the cost function, and they are also safe from the point of view of preventing catastrophic consequences for conditions that are far from nominal.

Deep RL frameworks and tools for mission planning and high-level decision-making for autonomous SC are considered in Ref. [21] under the assumption that subtasks are solved at the design stage accordingly. Two typical tasks, reflecting the problems of autonomous orbit insertion and the planning of scientific operations, are presented in the form of a partially observable MDP. The possibility of solving these problems using RL is considered, and the advantages, difficulties, and some features inherent to this approach are demonstrated. The dependence of the success of solving problems on the initial conditions and learning strategy is analyzed. The results of solving these problems demonstrate the possibility of using RL to improve or refine the policies obtained within the framework based on the paradigm focused on specific modes of operation while maintaining robustness to the uncertainty of environmental parameters.

RL methods are adapted to the paradigm of the SC finite state machine [21]. A Deep Q-learning algorithm is applied to partially observable MDP for obtaining policies that are comparable in performance to those that can be developed with prior knowledge of all the features of the problem. Various structures of the reward function, hyperparameters, and environment parameters were considered. The lack of positive results in solving the SC control tasks using deep Q-learning is a consequence of an insufficient amount of space domain data for training. This problem is aggravated by the strong computational requirements needed to run the environment simulator, which significantly slows down the learning task compared to simpler environments.

Moreover, the mode-based paradigm for the design of future decision-making algorithms is directly testable through the theory of hybrid systems. This paper presents one approach by which this theory can be used to identify "successful" or "stable" autonomous decision-making agents. Further work will investigate model-based RL methods to reduce the number of attempts and use existing knowledge of the space environment. In addition, fast models built using the Basilisk astrodynamic framework will be used to reduce training time.

CONCLUSION

Recent studies have shown the advantages of DL for solving space guidance, navigation, and control tasks. These results provide the basis for further studies of the possibilities of DL for controlling all or part of the SC on-board decision-making system.

Among the problematic issues that restrain the use of DL methods for the considered tasks, it should be noted, first of all, that the efficiency of solving problems is mainly illustrated by computer simulations, and there are practically no rigorous analytical results that provide stability and performance guarantees. Such results are very important for space practitioners for more active use of these methods in real missions. Examples of the efforts in this direction are Refs. [3, 25], where the methods of deep RL and classical control theory are used together to obtain stability guarantees.

As the next issue, it should be pointed out that many SC control tasks do not allow critical errors in the process of finding the optimal solution. In this regard, the ideas of such a direction as a safe-RL [39] should be more actively used for space missions.

The low training efficiency of DL algorithms is especially acute in space applications, which is due to the limited capabilities of SC for collecting and processing data in orbit. However, model-based RL methods [12] and transfer learning [37] has the potential to mitigate this problem.

Despite the underlined issues, the navigation, guidance, and control algorithms based on DL can simplify the development and increase the reliability of the SC control systems since the same algorithm can be used for a large number of different missions. DL makes it possible to develop control systems that can improve their performance using data accumulated during the operation of a particular object. This feature allows the designers to relax the requirements for the units of control systems (sensors, actuators), not to use special bench equipment, and reduce the development time and cost.

REFERENCES

1. Alpatov A., Khoroshylov S., Lapkhanov E. (2020). Synthesizing an algorithm to control the attitude motion of spacecraft equipped with an aeromagnetic deorbiting system. *East.-Eur. J. Enterprise Technol.*, **1**, № 5, 37—46.
2. *Artificial intelligence: a modern approach* (2010). Eds. S. J. Russell, P. Norvig. Pearson education. Inc. ISBN-13: 978-0134610993.
3. Berkenkamp F., Turchetta M., Schoellig A. P., Krause A. (2017). *Safe model-based reinforcement learning with stability guarantees*. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 908—919.
4. Biggs J., Fournier H. (2020). Neural-network-based optimal attitude control using four impulsive thrusters. *J. Guidance Control and Dynamics,* **43**, 299—309.
5. Chan D. M., Agha-Mohammadi A. (2019). *Autonomous imaging and mapping of small bodies using deep reinforcement learning*. IEEE Aerospace Conference, Big Sky, MT, USA, 2-9 March 2019, 1—12. DOI: 10.1109/AERO.2019.8742147.
6. Charles E. Oestreich, Linaresy R., Gondhalekarz R. (2021). Autonomous six-degree-of-freedom spacecraft docking maneuvers via reinforcement learning. *J. Aerospace Inform. Syst.*, **18,** № 7. DOI: 10.2514/1.I010914.
7. *Deep Learning* (2016). Eds. I. Goodfellow, Y. Bengio, A. Courville. The MIT Press. ISBN 978-0262035613.
8. Elkins J., Sood R., Rumpf C. (2020). *Autonomous spacecraft attitude control using deep reinforcement learning*. 71st International Astronautical Congress (IAC), The CyberSpace Edition, 12—14 October 2020.
9. Elkins J., Sood R., Rumpf C. (2020). *Adaptive continuous control of spacecraft attitude using deep reinforcement learning*. AAS/AIAA 2020 Astrodynamics Specialist Conference, 420—475.
10. Fujimoto S., Hoof H. V., Meger D. (2018). *Addressing function approximation error in actor-critic methods*. Int. Conf. on Machine Learning, PMLR, 1587—1596.
11. Furfaro R., Simo J., Gaudet B., Wibben D. R. (2013) *Neural-based trajectory shaping approach for terminal planetary pinpoint guidance*. AAS/AIAA Astrodynamics Specialist Conference, Hilton Head, SC, 1—18.
12. García J., Fernández F. (2015). A comprehensive survey on safe reinforcement learning. *J. Mach. Learn. Res.,* **16**, 1437—1480.
13. Gaudet B., Furfaro R. (2012). *Robust spacecraft hovering near small bodies in environments with unknown dynamics using reinforcement learning*. AIAA/AAS Astrodynamics Specialist Conference 2012, Minneapolis, Minnesota. DOI: 10.2514/6.2012-5072.
14. Gaudet B., Furfaro R. (2014). Adaptive pinpoint and fuel efficient Mars landing using reinforcement learning. *IEEE/CAA J. Automatica Sinica*, **1**, № 4, 397–411.
15. Gaudet B., Furfaro R., Linares R. (2020). *Reinforcement meta-learning for angle-only intercept guidance of maneuvering targets*. AIAA Scitech 2020 Forum. American Institute of Aeronautics and Astronautics Inc, AIAA, 1—16. DOI: 10.2514/6.2020-0609.
16. Gaudet B., Linares R., Furfaro R. (2020). Six degree-of-freedom hovering using LIDAR altimetry via reinforcement meta-learning. *Acta Astronautica*, **172**, 90–99.
17. Gaudet B., Linares R., Furfaro R. (2020). Seeker based adaptive guidance via reinforcement meta-learning applied to asteroid close proximity operations. *Acta Astronautica*, **171**, 1—13.
18. Gaudet B., Linares R., Furfaro R. (2020). Adaptive guidance and integrated navigation with reinforcement meta-learning. *Acta Astronautica*, **169**, 180—190.
19. Gers F. A., Schraudolph N. N., Schmidhuber J. (2002). Learning precise timing with LSTM recurrent networks. *J. Mach. Learn. Res.*, **3**, 115—143.
20. Giegerich R., Meyer C., Steffen P. (2004). A discipline of dynamic programming over sequence data. *Sci. Computer Program.*, **51**, № 3, 215—263.
21. Harris A., Teil T., Schaub H. (2019). Spacecraft decision-making autonomy using deep reinforcement learning. *Adv. Astronaut. Sci. AAS/AIAA Spaceflight Mechanics*, **168**, 1757—1776.
22. Hovell K., Ulrich S. (2020). *On deep reinforcement learning for spacecraft guidance*. AIAA SciTech Forum, 6—10 January 2020, Orlando, FL. DOI: 10.2514/6.2020-1600.
23. Huang Y., Li S., Sun J. (2018). Mars entry fault-tolerant control via neural network and structure adaptive model inversion. *Adv. Space Res.*, **63**, № 1, 557—571.
24. Izzo D., Märtens M., Pan B. (2019). A survey on artificial intelligence trends in spacecraft guidance dynamics and control. *Astrodyn.,* **3**, 287—299. DOI: 10.1007/s42064-018-0053-6.
25. Kamalapurkar R., Walters P., Dixon W. (2016). Model-based reinforcement learning for approximate optimal regulation. *Automatica*, **64**, 94—104.
26. Khoroshylov S. (2020). Relative control of an ion beam shepherd satellite in eccentric orbits. *Acta Astronautica*, 76, 89—98.

27. Khoroshylov S. V., Redka M. O. (2019). Relative control of an underactuated spacecraft using reinforcement learning. *Technical Mechanics*, **4**, 43—54.
28. Khoroshylov S. V., Redka M. O. (2019). *Control of a spacecraft using artificial intelligence. Cosmic technologies: present and future*. Dnipro, 63.
29. Khoroshylov S. V., Redka M. O. (2019). *Control of a spacecraft using reinforcement learning. Information technology in metallurgy and mechanical engineering*. Dnipro, 132 [in Ukrainian].
30. Khoroshylov S. V., Redka M. O. (2019). Intelligent spacecraft control using reinforcement learning. *Technical Mechanics*, **4**, 29—43 [in Ukrainian].
31. Koizumi S., Kikuya Y., Sasaki K., Masuda Y., Iwasaki Y., Watanabe K., Yatsu Y., Matunaga S. (2018). *Development of attitude sensor using deep learning*. AIAA/USU Conference on Small Satellites, AIAA, Session 7: Advanced Concepts II.
32. Krizhevsky A., Sutskever I., Hinton G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communs ACM*, **60** (6), 84—90. DOI: 10.1145/3065386.
33. Lafarge N. B., Miller D., Howell K. C., Linares R. (2020). *Guidance for closed-loop transfers using reinforcement learning with application to libration point orbits*. AIAA Guidance, Navigation, and Control Conference, Orlango, FL, 1—5.
34. Lary D. J. (2010). Artificial intelligence in Aerospace. *Aerospace Technologies Advancement, (INTECH), Croatia*. 492 p.
35. *Machine Learning* (1997). Ed. T. Mitchell. New York: McGraw Hill. ISBN 0070428077.
36. Ma Z., Wang Y., Yang Y., Wang Z., Tang L., Ackland S. (2018). Reinforcement learning-based satellite attitude stabilization method for non-cooperative target capturing. *Sensors*, **18**, № 12, 4331. DOI: 10.3390/s18124331.
37. Pan S. J., Yang Q. (2010). A survey on transfer learning. *IEEE Trans. Knowledge and Data Eng.*, **22**, № 10, 1345—1359. DOI: 10.1109/TKDE.2009.191.
38. Pierson H., Gashler M. (2017). Deep learning in robotics: a review of recent research. *Adv. Robotics*, **31**, № 16, 821—835. DOI: 10.1080/01691864.2017.1365009.
39. Polydoros A. S., Nalpantidis L. (2017). Survey of model-based reinforcement learning: applications on robotics. *Intel. Robot Syst.*, **86**, 153—173. DOI: 10.1007/s10846-017-0468-y.
40. Redka M. O., Khoroshylov S. V. (2019). *Application of reinforcement learning for spacecraft handling.* Computer modeling and optimization of complex systems, Dnipro, 150—151 [in Ukrainian].
41. *Reinforcement learning: an introduction* (1998). Eds. R. S. Sutton, A. G. Barto. MIT press. ISBN 978-0262193986.
42. Sallab A. E., Abdou M., Perot E., Yogamani S. (2017). Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, **19**, 70—76.
43. Sanchez-Sanchez C., Izzo D. (2018). Real-time optimal control via deep neural networks: study on landing problems. *J. Guidance, Control, and Dyn.*, **41**, № 5, 1122—1135.
44. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. (2017). Proximal policy optimization algorithms. *arXiv preprint*, arXiv:1707.06347.
45. Scorsoglio A., Furfaro R., Linares R., Gaudet B. (2020). *Image-based deep reinforcement learning for autonomous lunar landing*. AIAA SciTech Forum 2020, Orlando, FL. DOI: https://doi.org/10.2514/6.2020-1910.
46. Shirobokov M. G., Trofimov S. P. (2019). Formation control in low-Earth orbits by means of machine learning methods. *Keldysh Institute preprints*, **19**, 32.
47. Silver D., Schrittwieser J., Simonyan K. (2017). Mastering the game of Go without human knowledge. *Nature*, **550**, 354—359. DOI:10.1038/nature24270.
48. Vedant J. T. Allison M. W., Ghosh A. (2019). *Reinforcement learning for spacecraft attitude control*. Proceedings of the International Astronautical Congress, International Astronautical Federation, IAC-19-C1.5.2.
49. Vilalta R., Drissi Y. (2002). A perspective view and survey of meta-learning. *Artificial Intel. Rev.*, **18**, 77—95. DOI: 10.1023/A:1019956318069.
50. Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E. (2018). *Deep learning for computer vision: a brief review.* Computational Intelligence and Neuroscience. DOI: 10.1155/2018/7068349.
51. Wang Y., Ma Z., Yang Y., Wang Z., Tang L. (2019). *A new spacecraft attitude stabilization mechanism using deep reinforcement learning method*. 8th European Conference for Aeronautics and Space Sciences (EUCASS). DOI: 10.13009/EUCASS2019-33.
52. Watkins C. J. C. H., Dayan P. (1992). Q-learning. *Mach. Learn.*, **8**, 279—292. DOI: 10.1007/BF00992698.
53. Weiss K., Khoshgoftaar T.M., Wang D. (2016). A survey of transfer learning. *Big Data*, **3** (9). DOI: 10.1186/s40537-016-0043-6.
54. Willis S., Izzo D., Hennes D. (2016). Reinforcement learning for spacecraft maneuvering near small bodies. *Adv. Astronaut. Sci.*, **158**, 1351—1368.

55. Xu K., Wu F., Zhao J. (2018). Model-based deep reinforcement learning with heuristic search for satellite attitude control. *Ind. Robot: Int. J.*, **46** (3). DOI: 10.1108/IR-05-2018-0086.
56. Young T., Hazarika D., Poria S., Cambria E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intel. Magazine*, **3**, 55—75. DOI: 10.1109/mci.2018.2840738.
57. Zarchan P. (2012). Tactical and strategic missile guidance. *Amer. Inst. Aeronaut. and Astronaut., Inc.*, 18—21. DOI: 10.2514/4.868948.
58. ZhiBin Z., XinHong L., JiPing A., WanXin M., GuoHui Z. (2020). Model-free attitude control of spacecraft based on PID-guide TD3 Algorithm. *Int. J. Aerospace Eng.*, **2020**, 1—13. DOI: 10.1155/2020/8874619.
59. Zhu X. X., Tuia D., Mou L., Xia G. S., Zhang L., Xu F., Fraundorfer F. (2017). Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. and Remote Sensing Mag.*, **5**(4), 8—36.

Стаття надійшла до редакції *25.07.2021* *Received 25.07.2021*
Після доопрацювання *25.09.2021* *Revised 25.09.2021*
Прийнято до друку *11.10.2021* *Accepted 11.10.2021*

С. В. Хорошилов, пров. наук. співроб., д-р техн. наук, проф.
ORCID.org/0000-0001-7648-4791
E-mail: skh@ukr.net
М. О. Редька, аспірант
E-mail: mix5236@ukr.net

Інститут технічної механіки Національної академії наук України і Державного космічного агентства України
вул. Лешко-Попеля 15, Дніпро, Україна, 49005

ГЛИБОКЕ НАВЧАННЯ ДЛЯ НАВІГАЦІЇ, НАВЕДЕННЯ ТА КЕРУВАННЯ В КОСМОСІ

Успіхи глибокого навчання призвели до революції в області штучного інтелекту, демонструючи можливості створення автономних систем з високим рівнем розуміння середовища функціонування. Ці успіхи, а також нові завдання та вимоги в галузі освоєння космосу зумовили підвищений інтерес протягом останніх років до методів глибокого навчання серед працюючих в космічній сфері вчених і практиків. Метою цієї оглядової статті є аналіз останніх досягнень в галузі використання глибокого навчання для вирішення завдань навігації, наведення та керування в космосі. Розглянуто завдання керування кутовим і відносним рухом космічних апаратів при вирішенні як традиційних, так і нових завдань, таких як сервісні операції в космосі. Проаналізовано роботи, присвячені застосуванню цих методів для виконання операцій посадки і зависання при реалізації місій на Місяць, Марс і астероїди. Для вирішення таких завдань використовуються як методи навчання з вчителем, так і навчання з підкріпленням. Розглянуто використання різних архітектур штучних нейронних мереж, в тому числі згорткові та рекурентні. Аналізується можливість спільного використання глибокого навчання і методів теорії керування для підвищення ефективності вирішення розглянутих завдань. Виділено складності, що обмежують застосування розглянутих методів для космічних застосувань. Позначені необхідні напрямки досліджень для вирішення цих проблем.

Ключові слова: космічний апарат, глибоке навчання, навігація, наведення, керування, штучна нейронна мережа, навчання з підкріпленням, посадка, зависання.