ADAPTIVE TRAFFIC SIGNAL CONTROL BASED ON REINFORCEMENT LEARNING WITH CONSIDERATION OF VEHICLE AND PEDESTRIAN FLOWS

DOI: 10.36724/2072-8735-2026-20-5-57-66

Ksenia A. Polyantseva,
Moscow Technical University of Communications and Informatics, Moscow, Russia, ***k.a.poliantseva@mtuci.ru***

Mikhail G. Gorodnichev,
Moscow Technical University of Communications and Informatics, Moscow, Russia, ***m.g.gorodnichev@mtuci.ru***

Vladimir P. Kulagin,
Moscow Technical University of Communications and Informatics, Moscow, Russia, ***v.p.kulagin@mtuci.ru***

Manuscript received 02 March 2026;
Accepted 05 May 2026

Keywords: *adaptive traffic light control, reinforcement learning, Deep Q-Network, traffic flows, pedestrian flows, SUMO, intelligent transportation systems*

This paper addresses the problem of adaptive traffic signal control with joint consideration of vehicle and pedestrian flows. A Deep Q-Network-based method is proposed, in which the agent decides whether to keep the current traffic signal phase or switch to the next one based on the current state of the intersection. This state includes traffic lane congestion, the number of pedestrians in the crossing zone, and the current control phase. The main contribution of the paper is the development of a multi-component adaptive reward function. It includes vehicle delays, pedestrian waiting times, intersection capacity, and a penalty for excessive phase switching. When pedestrians accumulate in the crossing area, the pedestrian-related component of the reward is increased. The experimental evaluation was carried out in the SUMO microscopic traffic simulation environment. The proposed method was compared with fixed traffic light control. The results showed a reduction in average pedestrian and vehicle wait times, shorter queue lengths, and an increase in intersection capacity. These results confirm that reinforcement learning combined with an adaptive reward function can be used for more flexible and balanced urban intersection management.

Information about authors:

Ksenia A. Polyantseva, *Associate Professor of the Department of Data Mining, Candidate of Technical Sciences, Moscow Technical University of Communications and Informatics, Moscow, Russia. ORCID: 0000-0002-7102-4208*

Mikhail G. Gorodnichev, *Head of the Department of Mathematical Cybernetics and Information Technology, Candidate of Technical Sciences, Moscow Technical University of Communications and Informatics, Moscow, Russia. ORCID: 0000-0003-1739-9831*

Vladimir P. Kulagin, *Professor of the Department of Mathematical Cyber-netics and Information Technology, Doctor of Technical Sciences, Moscow Technical University of Communications and Informatics, Moscow, Russia. ORCID: 0000-0001-8826-6724*

Для цитирования:

Полянцева К.А., Городничев М.Г., Кулагин В.П. Адаптивное управление светофорами на основе обучения с подкреплением с учетом потоков транспортных средств и пешеходов // T-Comm: Телекоммуникации и транспорт. 2026. Том 20. №5. С. 57-66.

For citation:

K.A. Polyantseva, M.G. Gorodnichev, V.P. Kulagin, "Adaptive traffic signal control based on reinforcement learning with consideration of vehicle and pedestrian flows," *T-Comm*, 2026, vol. 20, no. 5, pp. 57-66.

Introduction

With increasing motorization and growing pressure on urban transport infrastructure, improving traffic management efficiency has become an important research and practical problem.

Traditional traffic signal control systems, based on fixed phase switching cycles, do not take into account the current traffic situation, leading to increased waiting times for vehicles and pedestrians, longer queues, and reduced intersection capacity.

Modern intelligent transport systems actively utilize machine learning and computer vision methods to analyze traffic situations in real time [1]. Of particular interest is the use of reinforcement learning, in which a control agent forms a control strategy based on interactions with the environment. Unlike predefined algorithms, this approach allows for adaptive traffic flow, accounting for queue accumulation, and minimizing overall time losses [2].

Despite a significant amount of research in the field of adaptive intersection control, most existing work focuses primarily on optimizing traffic flows without explicitly considering pedestrian traffic. This limits the applicability of such solutions in urban environments, where the interests of all road users must be taken into account.

This raises the challenge of developing management methods that would jointly optimize traffic and pedestrian flows within a single model. This approach not only increases intersection capacity but also ensures more balanced management of various road users.

The aim of this paper is to develop and evaluate a method for adaptive traffic light control based on reinforcement learning, taking into account vehicle and pedestrian flows. To achieve this goal, the following tasks are addressed:

- developing a state model of a signalized intersection;
- developing a reward function that takes into account the delays of various traffic participants;
- constructing an algorithm for selecting control actions;
- experimentally evaluating the effectiveness of the proposed approach.

The main contribution of this paper lies not in the application of DQN to the traffic light control problem itself, but in the construction of a learning scheme with a multi-component adaptive reward function. Unlike vehicle-centric approaches, the proposed function explicitly accounts for pedestrian delays, dynamically enhances the pedestrian component as people accumulate in the crossing zone, and additionally penalizes excessive phase switching. This allows the agent to learn not only to increase throughput but also to maintain a balance between vehicle and pedestrian flows.

The practical significance of this paper lies in the potential application of the proposed approach to the development of intelligent traffic management systems within the framework of smart city concepts [3].

1. Related work

The development of traffic signal control methods has evolved from rigid fixed cycles to adaptive systems that use data on the current state of the transport network [4]. Classic approaches rely on predetermined phase plans and work well only under conditions of relatively stable demand. When traffic volume fluctuates unevenly, such schemes quickly lose effectiveness: green time may be allocated to a lightly loaded route, while queues form on other routes. This is why intelligent transport systems are increasingly focusing on methods that can adjust control based on the observed traffic situation.

Adaptive control systems, such as SCATS and SCOOT, have become important steps in this direction. They use data from road sensors and allow traffic signal control parameters to be modified based on the current traffic load. These solutions have demonstrated that the transition from fixed cycles to adaptive control does indeed reduce delays and improve throughput. However, such systems are typically based on heuristic rules, statistical models, or local optimization. They require manual tuning and are not always well-suited to new traffic scenarios. There's a lack of flexibility.

Reinforcement learning methods have seen significant development in recent years. In this approach, a traffic light is considered an agent that observes the state of the intersection, chooses an action, and receives a reward for its performance [5]. This approach eliminates the need to manually specify a switching policy, instead deriving it through interaction with the environment. Value-based methods, particularly DQN and its modifications, were most frequently used in early studies. They are suitable for problems with a discrete set of actions, such as choosing between maintaining the current phase and switching to the next. However, the performance of such methods depends heavily on how the state and reward function are defined. If the state reflects one aspect of the traffic flow, while the reward optimizes another, learning becomes less robust.

Modern research is gradually moving away from simple single-node approaches. For single intersections, Double DQN, Dueling DQN, PPO, and other deep reinforcement learning approaches are increasingly being used to improve training stability and accelerate convergence. For example, several studies have shown that improving the agent architecture, using a prioritized experience buffer, or switching to actor-critic methods can reduce average waiting times and queue lengths compared to fixed-time control and classical adaptive heuristics. However, such solutions are primarily focused on vehicle traffic flow. Pedestrians are either not considered or are only indirectly included.

A separate line of research concerns intersection network management. Here, the problem becomes more complex: the decision of one traffic light affects neighboring nodes, and a locally optimal action can worsen the condition of the entire traffic corridor. To address this problem, multi-agent reinforcement learning, graph neural networks, and inter-agent communication mechanisms are being used. In such approaches, each intersection can be treated as a separate agent, and information exchange between agents allows for the propagation of queues and the interaction of phases to be taken into account. Graph models additionally provide the ability to describe the network topology and the differences between intersections. This is an important step forward, especially for urban networks. However, the price of such flexibility is high: architectures become more complex, the need for communication between nodes increases, and transferring the model to another network requires additional testing.

Another area of research concerns the portability and generalizability of models. Federated learning, transfer learning, and methods that allow for training control systems for more than just one specific intersection configuration are increasingly discussed in the literature. This stems from a practical problem: a model that performs well in simulation for one network may perform poorly when the geometry, traffic volume, or phase set changes. Federated and graph approaches partially address this problem by allowing knowledge exchange between nodes or regions without transferring raw data. However, most of these solutions are still primarily tested in simulators. Real-world implementation remains rare.

Research that considers not only cars but also pedestrians is particularly important for this work. In an urban environment, traffic signal control cannot be viewed solely as a task of minimizing traffic delays. If only vehicle flow is optimized, the system may generate a strategy that is convenient for cars but uncomfortable or unsafe for pedestrians. Some studies propose pedestrian-aware approaches, where the objective function includes the total user delay or the safety indicators of pedestrian-vehicle conflicts. These studies demonstrate that explicitly considering the pedestrian component changes the nature of control and allows for more balanced decisions. However, such studies are significantly fewer in number than vehicle-centric approaches. This is a significant gap in the field.

Integrating computer vision also remains an important, but not entirely resolved, problem. Most reinforcement learning algorithms in traffic signal control use data obtained from a simulator or from idealized sensors: queue lengths, speeds, lane occupancy, and the number of vehicles. In a real urban environment, such data is often obtained from cameras, requiring a link between the perception and decision-making modules. There are studies where control is based on visual data or end-to-end processing of video streams. However, this area is still less mature than research based on features directly extracted from SUMO or detector data. Computer vision is especially rarely combined with explicit consideration of pedestrian flows in the reward function.

A general comparison of the approaches considered is provided in Table 1. While the table does not claim to be a comprehensive coverage of the literature on the topic, it highlights the key point: while most modern approaches have developed the algorithmic component of control well, the pedestrian component and visual perception data are still used infrequently.

Table 1

Comparison of approaches to adaptive control of traffic lights

Source	**Method**	**Scale**	**Pedestrian consideration**	**Main limitation**
[6]	Double DQN	1 intersection	No	Focus on transport metrics
[7]	PN-D3QN	1 intersection	No	No network coordination
[8]	PPO	1 intersection	No	Pedestrian flows are not included in the model
[9, 10]	MARL, agent communication	Network of intersections	No	Increasing complexity of architecture and data exchange
[11-13]	Graph MARL	Network of intersections	No	High complexity of the graph model
[14, 15]	Federated RL	Several networks	No	Testing primarily in simulation
[16]	Pedestrian-aware DRL	1 intersection	Yes	Limited number of scenarios and topologies
[17]	Safety-oriented DRL	1 intersection	Yes	A developed infrastructure of connected vehicles is required
[18, 19]	Vision-based control	1 intersection	Partially	The complexity of the transition from simulation to a real environment
The proposed approach	DQN with multi-component reward function	1 intersection	Yes	Requires further expansion to the intersection network

The comparison shows that reinforcement learning has already become one of the main tools for adaptive traffic signal control [20]. The most advanced approaches are related to multi-agent coordination, graph models, policy portability, and multi-objective optimization [21].

However, an important limitation remains: most work still focuses on vehicular metrics [22]. Pedestrian flows are considered less frequently, and the integration of computer vision data into the decision-making loop is usually considered separately from the problem of balancing vehicle and pedestrian delays.

The proposed method uses reinforcement learning to control a traffic signal, but the reward function is formed based on more than just vehicle metrics. It includes pedestrian delays, throughput, and penalties for excessive phase changes. This approach allows intersection control to be viewed as a problem of balancing multiple types of traffic participants, rather than as optimizing a single vehicle flow.

2. Problem statement

We consider a signalized intersection where vehicle and pedestrian traffic fluctuate over time. Traffic signal control is performed discretely: at each simulation step, the system receives information about the current state of the intersection and makes a decision to maintain or change the current phase. This setup aligns well with reinforcement learning problems, where the control agent sequentially interacts with the environment and adjusts its strategy based on the results of previous actions.

The problem is formalized as a Markov decision process:

$$\langle S, A, P, R \rangle,$$

where S is the set of environment states, A is the set of admissible actions, P is the state transition function, and R is the reward function. In this paper, the transition function is not defined analytically: it is determined by the dynamics of the transport environment in SUMO. The agent observes only the current state, applies an action, and receives a new state along with the reward value.

The state of the environment at time t is defined by the vector

$$s_t = \{n_1, n_2, \dots, n_{12}, p, \phi\},$$

where n_i is the number of vehicles in the i-th traffic lane, p is the number of pedestrians in the crossing zone, and ϕ is the current traffic light phase. Thus, the agent obtains a compact description of the intersection: lane occupancy, pedestrian activity, and the current traffic light mode. The state vector has 14 dimensions.

Before feeding the values of n_i and p into the neural network model, they are normalized. This is necessary not only to speed up training. With different scales of input features, the network may overestimate some state components and largely ignore others. Normalization reduces this effect and makes learning more robust.

The agent's action set is defined as

$$A = \{a_0, a_1\},$$

where a_0 means maintaining the current phase, and a_1 switching to the next phase. This discrete set of actions was chosen deliberately. It simplifies control and makes the problem suitable for DQN, while still preserving the core control logic: the agent decides whether the current phase is sufficient to remain active or whether the traffic situation already requires switching.

In a real system, a traffic light cannot change phases arbitrarily frequently. This would create unstable control and could worsen traffic safety. Therefore, a minimum phase duration constraint is introduced. If a phase is active for less than the specified time, the switching action is blocked or considered undesirable. Additionally, a penalty for excessive switching is included in the reward function.

The reward function must reflect several goals simultaneously [23]. It is necessary to reduce vehicle waiting times, prevent excessive pedestrian waiting times, maintain intersection capacity, and avoid excessively frequent phase changes. The following reward form is used in this work:

$$R_t = -\alpha W_c - \beta W_p + \eta Q + \delta P_{prio} - \lambda S_{switch},$$

where W_c is the total waiting time of vehicles, W_p is the total waiting time of pedestrians, Q is the number of objects that passed the intersection per simulation step, P_{prio} is the pedestrian flow priority coefficient, and S_{switch} is the phase switching penalty. The coefficients α, β, η, δ and λ determine the relative importance of individual components.

Pedestrian priority is introduced not as a constant bonus, but as a response to the accumulation of pedestrian flow:

$$P_{prio} = \begin{cases} 1, & p < p_{th}, \\ k, & p \geq p_{th}, \end{cases}$$

where p_{th} is the pedestrian threshold, and $k > 1$ is the priority gain coefficient. This scheme avoids giving excessive priority to the pedestrian phase when the number of people is small, but increases its influence when the wait at the crossing becomes noticeable. It's a simple mechanism, but it makes the control less car-centric.

Additionally, the β coefficient, which is responsible for the penalty for waiting for pedestrians, can be adjusted. If the number of pedestrians exceeds the threshold, the weight of the pedestrian delay increases:

$$\beta_t = \begin{cases} \beta, & p < p_{th}, \\ \beta \cdot k_\beta, & p \geq p_{th}, \end{cases}$$

where $k_\beta > 1$ is the gain. In this case, the reward function becomes adaptive: it changes the balance of objectives depending on the current situation at the intersection.

The agent's goal is to find a strategy π that maximizes the expected discounted reward sum:

$$\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right],$$

where γ is the discount factor. The optimal strategy should not minimize a single metric, but rather balance between vehicle delays, pedestrian waiting times, throughput, and phase switching stability.

The Deep Q-Network algorithm is used to solve the problem. It approximates the action value function

$$Q(s, a; \theta),$$

where θ are the neural network parameters. The choice of DQN is due to the fact that the action space in the considered setting is discrete, and the state has a fixed dimension. This allows for the use of a relatively simple agent architecture and its training directly through interaction with the simulated environment.

3. Proposed method

This paper proposes a method for adaptive traffic light control based on the Deep Q-Network algorithm. Its core idea is to replace the predefined phase switching logic with a strategy that evolves as the agent interacts with the traffic environment. The agent receives the current intersection state, evaluates possible actions, and selects one. The environment then transitions to a new state, and the agent receives the value of the reward function. This gradually forms a control policy.

The overall design of the method includes several interconnected components: a simulation environment, a data collection module, a state generation module, a DQN agent, a control action application mechanism, and a reward calculation unit. SUMO is used as the traffic environment, and data exchange between the simulator and the control algorithm is accomplished via the TraCI interface [24]. This separation is convenient: SUMO is responsible for traffic dynamics, while the Python control module implements the learning algorithm and action selection.

At each simulation step, the system reads data on the number of vehicles in the traffic lanes and the number of pedestrians in the crossing zone. These values are combined with the current traffic light phase to form a state vector, s_t. The vector is then fed to a neural network model, which returns estimates, $Q(s_t, a)$, for the permissible actions. If the agent selects the phase-maintaining action, the traffic light continues to operate in its current mode. If the phase-switching action is selected, a phase-change command is sent via TraCI. A minimum phase duration constraint is also checked to prevent excessively frequent phase-switching.

The DQN agent architecture is compact. The input layer has a dimension of 14, corresponding to the number of features in the state vector. Next, two fully connected hidden layers with 64 neurons each and a ReLU activation function are used. The output layer contains two neurons, each corresponding to one agent action: a_0 maintain the current phase, a_1 switch the phase.

Action selection in operational mode is performed according to the rule

$$a_t = \arg\max_{a \in A} Q(s_t, a; \theta),$$

where θ are the parameters of the underlying neural network.

During training, a ε-greedy strategy is used. With probability ε, the agent chooses a random action, and with probability $1-\varepsilon$, the action with the maximum value of the *Q*-function. At the beginning of training, a large ε value allows the agent to explore different behavioral options. Then, it gradually decreases, and the agent increasingly uses its accumulated experience. This is a simple mechanism, but without it, the model quickly gets stuck in a random local strategy.

To improve the stability of training, an experience replay buffer is used. Transitions of the form

$$(s_t, a_t, R_t, s_{t+1}, d_t),$$

where d_t is the end-of-episode indicator. Training is performed not on consecutive observations, but on random mini-samples from the buffer. This reduces the correlation between adjacent states, and the neural network receives more diverse examples. This is important for a transport environment: states in it often change smoothly, and training only on adjacent steps makes the model less stable.

In addition to the main network, a target network with parameters θ^- is used. It is updated less frequently than the main model and is used to calculate target values. For a transition from the buffer, the target value is specified as

$$y_t = \begin{cases} R_t, & d_t = 1, \\ R_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-), & d_t = 0. \end{cases}$$

The main network parameters are then updated by minimizing the error between the current estimate $Q(s_t, a_t; \theta)$ and the target value y_t. The Huber loss is used as the loss function, as it is less sensitive to outliers than the mean squared error.

The DQN agent training process can be described by the following sequence of steps:

1. The SUMO environment and the TraCI interface are initialized.
2. The main network $Q(s, a; \theta)$ and the target network $Q(s, a; \theta^-)$ are initialized.
3. A replay buffer is created.
4. At the beginning of each episode, the environment is initialized to s_0.
5. At each simulation step, the agent selects an action a_t using the ε-greedy strategy.
6. The selected action is applied to the traffic light object via TraCI.
7. The environment transitions to a new state s_{t+1}, after which the reward R_t is calculated.
8. The transition $(s_t, a_t, R_t, s_{t+1}, d_t)$ is saved in the experience buffer.
9. Based on a random mini-sample from the buffer, the main network parameters are updated.
10. After a specified number of steps, the target network is updated: $\theta^- \leftarrow \theta$.
11. At the end of the episode, the value of ε is decreased.

The unique feature of the proposed method lies not only in the use of DQN, but also in the way the training signal is generated. The reward function takes into account vehicle wait times, pedestrian wait times, throughput, and penalties for excessive switching. Therefore, the agent is not trained to maximize vehicle flow alone. It is forced to find a compromise. Sometimes, it is more advantageous to maintain the current phase to relieve the accumulated vehicle queue. In other situations, it is better to switch earlier and prevent excessive pedestrian wait times. This is precisely the logic required for an urban intersection.

An additional advantage of the method is that it allows for the integration of different data sources. In a simulated environment, vehicle information can come from SUMO detectors, while in a more practical setting, a similar state vector can be generated based on camera data or computer vision models [25]. This allows the proposed approach to be considered as a basis for further transition from simulation testing to more realistic intelligent traffic management scenarios.

Thus, the proposed scheme differs from the basic DQN approach in the structure of the training signal. The agent is rewarded not only for reducing transport delays but also for reducing pedestrian wait times and for maintaining the stability of the phase control. When the pedestrian count exceeds a specified threshold, the contribution of pedestrian delays increases, allowing the control to adapt to local pedestrian flow accumulation.

4. Experimental part

The proposed method was tested using the SUMO microscopic modeling environment. A signalized intersection with four traffic directions was constructed. Each direction included three lanes, allowing for separate consideration of the traffic load on the approaches to the intersection. In addition to vehicles, pedestrian flows crossing the roadway at signalized crossings were added to the model.

Traffic signal control was performed via the TraCI interface. At each simulation step, the control module received data on the intersection state, generated an input vector for the DQN agent, and applied the selected action to the traffic signal. The state vector included the number of vehicles in the 12 lanes, the number of pedestrians in the crossing zone, and the current traffic signal phase. Thus, the agent made decisions based on the current traffic situation rather than on a predetermined schedule.

Fixed control was used as the baseline method. In this mode, phases switched according to a predetermined cycle and were independent of the observed number of vehicles or pedestrians. The comparison was conducted using the same network topology, the same traffic routes, and the same simulation duration. This allows us to evaluate the impact of the control algorithm, rather than differences in the experimental conditions.

The main parameters of the experimental setup are presented in Table 2. The parameter values must be the same for all compared runs.

Table 2

Parameters of the experimental setup

Parameter	Value
Simulation environment	SUMO
Network type	One controlled intersection
Number of directions	4
Number of lanes	12
Pedestrian Crossings	Controlled Crossings at Intersections
Phase diagram	Two main phases: movement in one direction and switching to the next phase
Minimum phase duration	10 s
Duration of one episode	1000 simulation steps
Simulation step	1 s
Number of training episodes	100
Number of experiment repetitions	5
Load Scenarios	Low, Medium, High Intensity
Basic Method	Fixed Traffic Light Cycle
Comparison method	DQN control with multi-component reward

The parameters of the DQN agent are given in Table 3. Their specification is important for reproducibility, since the quality of training significantly depends on the buffer size, learning rate, discount factor, and ε reduction strategy.

Table 3

DQN agent parameters

Parameter	Value
Dimension of the input state	14
Number of actions	2
Hidden layers	2 layers of 64 neurons each
Activation function	ReLU
Optimizer	Adam
Learning rate	0.001
Discount factor γ	0.95
Replay buffer size	10000 transitions
The size of the mini-sample (batch size)	32
Minimum buffer size before training	1000 transitions
Loss function	Huber loss
Initial value ε	1.0
Minimum value ε	0.01
The rate of decrease ε	0.005
Target network update frequency	every 10 episodes

Routes were generated using standard SUMO tools. The model included cars, buses, trucks, and pedestrians. To test the algorithm's robustness, scenarios with varying traffic volumes were considered: low, medium, and high. In the baseline experiment, the results were aggregated across all runs, allowing for an averaged assessment of the method's effectiveness.

Four metrics were used to evaluate traffic quality:

- average vehicle waiting time;
- average pedestrian waiting time;
- total vehicle queue length;
- intersection throughput.

These metrics were chosen because they reflect different aspects of traffic signal operation. Vehicle waiting time characterizes the efficiency of vehicle traffic. Pedestrian waiting time indicates whether gains are achieved at the expense of poor conditions for pedestrians. Queue length reflects the accumulation of congestion, while throughput measures how many vehicles pass through the intersection per unit of time.

The results of the comparison of fixed and adaptive control are presented in Table 4. The values are presented as averages over a series of independent runs with standard deviations. This format allows us to evaluate not only the average gain of the adaptive control but also the robustness of the results.

Table 4

Comparison of fixed and adaptive control, mean ± standard deviation

Metric	Fixed control	Adaptive control
Average waiting time for pedestrians, steps	45.0 ± 2.1	25.0 ± 1.6
Average waiting time for transport, steps	60.0 ± 2.8	42.0 ± 2.0
Total queue length, m	180.0 ± 7.4	120.0 ± 5.6
Throughput, veh/min	50.0 ± 1.9	65.0 ± 2.3

Additionally, the method was tested at different traffic levels. This breakdown is important, as a strategy effective at moderate traffic flow may perform less well at a saturated intersection. Table 5 shows the results for low, medium, and high traffic volumes.

Table 5

Comparison of control methods at different load levels, mean ± standard deviation

Scenario	Method	Pedestrian waiting, steps	Vehicle waiting, steps	Queue, m	Throughput, veh/min
Low load	DQN	31.0 ± 1.5	39.0 ± 2.0	96.0 ± 4.8	38.0 ± 1.4
Low load	Fix	20.0 ± 1.2	30.0 ± 1.7	72.0 ± 3.9	44.0 ± 1.6
Average load	DQN	45.0 ± 2.1	60.0 ± 2.8	180.0 ± 7.4	50.0 ± 1.9
Average load	Fix	25.0 ± 1.6	42.0 ± 2.0	120.0 ± 5.6	65.0 ± 2.3
High load	DQN	62.0 ± 3.4	84.0 ± 4.6	255.0 ± 11.2	56.0 ± 2.4
High load	Fix	38.0 ± 2.7	61.0 ± 3.5	178.0 ± 8.9	72.0 ± 2.8

Table 5 shows that as load increases, the absolute values of delays and queue lengths increase, but the adaptive scheme maintains control over the intersection. The effect is most noticeable in medium- and high-load scenarios, where the fixed cycle leads to queue accumulation more quickly.

For a visual comparison of these same indicators, the diagram shown in Figure 1 is shown. It demonstrates that adaptive control reduces the values of metrics related to waiting and queues while simultaneously increasing throughput.

Table 4 and Figure 1 show that adaptive control improves all the metrics analyzed. Average pedestrian wait time is reduced from 45 to 25 steps, a reduction of approximately 44%. Vehicle wait time is reduced from 60 to 42 steps, a 30% reduction. The total queue length is reduced from 180 to 120 meters. Throughput increases from 50 to 65 vehicles per minute.

To evaluate the learning process, the total reward dynamics were plotted across episodes. This graph allows us to verify whether the agent is developing a stable control strategy, rather than simply experiencing random improvements in individual runs.

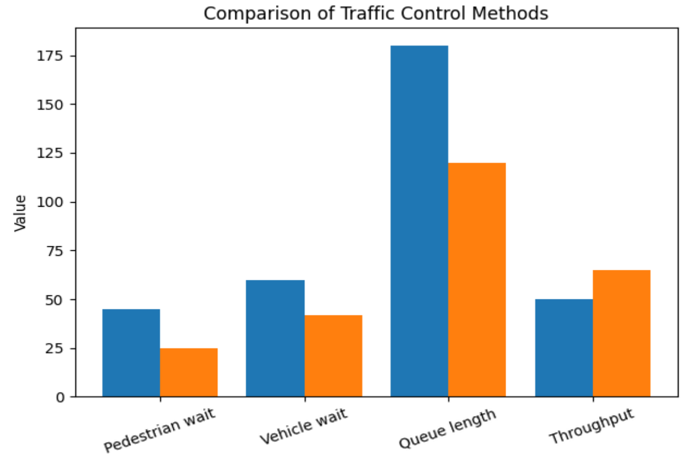

Fig. 1. Comparison of the performance indicators of fixed and adaptive control

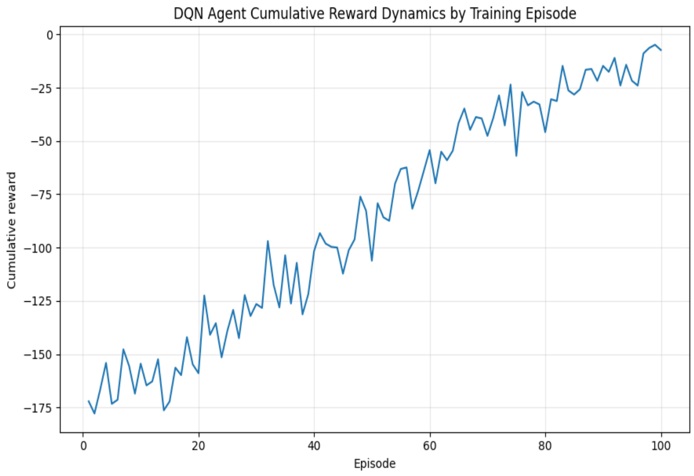

Fig. 2. Dynamics of the total reward of the DQN agent by training episodes

As can be seen from Figure 2, the reward value changes unstably in the initial episodes, which is due to the high proportion of random actions in the ε-greedy strategy. As ε decreases, the agent more frequently utilizes its accumulated experience, and the total reward gradually stabilizes. This indicates the development of a more stable control policy.

The obtained values should be considered the result of an initial simulation test. They demonstrate the viability of the proposed approach but do not constitute exhaustive proof of its superiority under all conditions. For a more rigorous evaluation, it is necessary to expand the experiment: add several load scenarios, compare the method not only with a fixed cycle but also with other adaptive or RL-based algorithms, and also report the spread of results across independent runs.

This result is due to the fact that the agent does not follow a fixed cycle but reacts to the current state of the intersection. If a queue forms in one direction, the strategy can hold the corresponding phase longer. If pedestrians accumulate at a crossing, the reward function increases the weight of pedestrian waiting time, and the agent is incentivized to give a green light more quickly. This improves not only traffic but also pedestrian performance.

5. Analysis of results

The results show that adaptive control reduces delays for both vehicles and pedestrians. This is an important point: the improvement is achieved not by redistributing waiting time from one group of road users to another, but by more flexible use of traffic light phases.

A fixed cycle predetermines the signal duration and does not react to actual queue buildup. A DQN agent, in contrast, uses the current state of the intersection and can change its behavior depending on the situation.

The standard deviation in Table 4 allows us to assess the robustness of the method. If the spread of values for adaptive control remains moderate, we can conclude not only that the average result has improved but also that the effect is reproducible. This is especially important for reinforcement learning problems, where the final policy may depend on the initial network initialization, the order of observations, and random actions during the exploration phase.

Dividing the experiments by load scenarios reveals the conditions under which adaptive control yields the greatest gains. At low traffic loads, the difference between the methods is usually smaller, since the fixed cycle does not yet introduce significant losses. At medium and high loads, the advantage of the DQN agent becomes more noticeable: the system encounters queues and pedestrians more frequently, providing more opportunities for adaptive phase reallocation.

The most noticeable improvement is observed in pedestrian waiting times. This is due to the fact that pedestrian flow is explicitly included in the environment state and the reward function. As the number of pedestrians in the crossing zone increases, the contribution of pedestrian delay to the overall learning signal increases. As a result, it becomes unprofitable for the agent to hold the vehicle phase for a long time if a group of people has already formed at the crossing. Simple logic, but for an urban intersection, it is crucial.

The reduction in vehicle waiting times is explained by another aspect of the same adaptivity. The agent takes lane congestion into account and can maintain the current phase if a significant flow of vehicles is passing through it. In a fixed-time control system, this is not possible: switching occurs according to a schedule, even if one direction is already empty and the other is congested. Therefore, some of the green time is wasted. In the adaptive mode, it is used more efficiently.

The reduction in the total queue length is due to the agent not only reacting to the accumulated delay but also gradually learning to avoid states in which the queue begins to grow rapidly. The experience replay buffer allows for the use of previous situations when updating the model, and the target network stabilizes learning. Therefore, the strategy becomes not simply reactive, but more stable over time.

The penalty for phase switching is particularly noteworthy. Without such a restriction, the agent might choose excessively frequent signal changes, attempting to instantly react to every small flow change. In practice, this is bad behavior: it creates instability, increases time losses during transition phases, and can degrade safety. Adding a penalty makes the policy more relaxed. The agent switches phases only when it provides a significant benefit to the overall reward function.

However, the results should not be overestimated. The experiment was conducted for a single signalized intersection, so it cannot yet be concluded that the same policy, without modification, would be effective for a network of intersections. In a multi-intersection system, additional effects arise: queue propagation between nodes, the need for phase coordination, the influence of adjacent traffic lights, and the heterogeneity of the road topology. This is a separate issue.

Another limitation is related to the quality of the input data. In the simulation, the agent receives fairly accurate information on the number of vehicles and pedestrians. In a real environment, this data may be obtained from cameras or other sensors, meaning it will contain detection errors, omissions, false alarms, and transmission delays. Such uncertainty can impact the quality of control. Therefore, before practical application, the method needs to be further tested under noisy observation conditions.

Despite these limitations, the experiment confirms the viability of the proposed approach. The multi-component reward function allows the agent to consider not just a single metric, but several related objectives: vehicle delays, pedestrian waiting times, throughput, and phase switching stability. This makes the system's behavior more balanced than with fixed-priority control.

Conclusion

This paper examines the problem of adaptive traffic signal control, simultaneously accounting for vehicle and pedestrian flows. A reinforcement learning-based method is proposed, in which the control agent uses intersection conditions, including lane occupancy, pedestrian count, and the current traffic signal phase. A Deep Q-Network algorithm is used to select control actions.

The key element of the method is a multi-component reward function. It includes vehicle delays, pedestrian waiting times, intersection capacity, and a penalty for excessive phase changes. This formulation allows the agent to not simply reduce vehicle queues but also to seek a more balanced control strategy. This is important for urban environments, where the effectiveness of traffic signal control cannot be assessed solely by transport metrics.

Experimental testing in the SUMO environment demonstrated the advantages of adaptive control over fixed-cycle control. The average pedestrian wait time decreased from 45 to 25 steps, vehicle wait time from 60 to 42 steps, the total queue length decreased from 180 to 120 meters, and throughput increased from 50 to 65 vehicles per minute. These results demonstrate that incorporating the pedestrian component into the reward function does influence agent behavior and helps avoid a vehicle-centric control strategy.

However, the work has limitations. The experiments were conducted for a single signalized intersection in a simulation environment, so further testing should include more complex road networks, multiple interacting traffic signals, and scenarios with noisy input data. A promising direction is to expand the method to multi-agent settings, as well as integrate computer vision data to generate environmental state in conditions closer to real urban infrastructure.

References

[1] K. A. Polyantseva, M. G. Gorodnichev, M. S. Moseva, and T. D. Potapchenko, "On the Applicability of Neural Networks in the Tasks of Detecting Dangerous Movement," in *Proc. Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russia, 2019, p. 8840654, doi: 10.1109/WECONF.2019.8840654.

[2] M. G. Gorodnichev, K. A. Polyantseva, and M. V. Yashina, "Informational Aspects of Automated Monitoring of Dangerous Driving," in *Proc. Systems of Signals Generating and Processing in the Field of on Board Communications (SOSG)*, Moscow, Russia, 2019, p. 8706721, doi: 10.1109/SOSG.2019.8706721.

[3] K. Polyantseva, M. Gorodnichev, Y. Mitrokhin, and M. Moseva, "On the Problem of Developing a Realistic Road Infrastructure Simulator for Reinforced Learning," in *Proc. Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russian Federation, 2023, pp. 1-8, doi: 10.1109/WECONF57201.2023.10147983.

[4] M. Noaeen, A. Naik, L. Goodman, J. Crebo, T. Abrar, Z. S. H. Abad, A. L. C. Bazzan, and B. Far, "Reinforcement learning in urban network traffic signal control: A systematic literature review," *Expert Systems with Applications*, vol. 199, p. 116830, 2022, doi: 10.1016/j.eswa.2022.116830.

[5] H. Zhao, C. Dong, J. Cao, and Q. Chen, "A survey on deep reinforcement learning approaches for traffic signal control," *Engineering Applications of Artificial Intelligence*, vol. 133, Part A, p. 108100, 2024, doi: 10.1016/j.engappai.2024.108100.

[6] S. Bouktif, A. Cheniki, A. Ouni, and H. El-Sayed, "Deep reinforcement learning for traffic signal control with consistent state and reward design approach," *Knowledge-Based Systems*, vol. 267, p. 110440, 2023, doi: 10.1016/j.knosys.2023.110440.

[7] C. Cai and M. Wei, "Adaptive urban traffic signal control based on enhanced deep reinforcement learning," *Scientific Reports*, vol. 14, p. 14116, 2024, doi: 10.1038/s41598-024-64885-w.

[8] L. Wang, G. Zhang, Q. Yang, and T. Han, "An adaptive traffic signal control scheme with Proximal Policy Optimization based on deep reinforcement learning for a single intersection," *Engineering Applications of Artificial Intelligence*, vol. 149, p. 110440, 2025, doi: 10.1016/j.engappai.2025.110440.

[9] Q. Jiang, M. Qin, S. Shi, W. Sun, and B. Zheng, "Multi-agent reinforcement learning for traffic signal control through universal communication method," in *Proc. 31st International Joint Conference on Artificial Intelligence (IJCAI-ECAI)*, Vienna, Austria, 2022, pp. 3854-3860.

[10] X. Song, B. Zhou, and D. Ma, "Cooperative Traffic Signal Control Through A Counterfactual Multi-Agent Deep Actor Critic Approach," *SSRN*, 2022, doi: 10.2139/ssrn.4021959.

[11] S. Yang, "Hierarchical graph multi-agent reinforcement learning for traffic signal control," *Information Sciences*, vol. 634, pp. 55-72, 2023, doi: 10.1016/j.ins.2023.03.087.

[12] Y. Bie, Y. Ji, and D. Ma, "Multi-agent Deep Reinforcement Learning collaborative Traffic Signal Control method considering intersection heterogeneity," *Transportation Research Part C: Emerging Technologies*, vol. 164, p. 104663, 2024, doi: 10.1016/j.trc.2024.104663.

[13] T. Wang, Z. Zhu, J. Zhang, J. Tian, and W. Zhang, "A large-scale traffic signal control algorithm based on multi-layer graph deep reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 162, p. 104582, 2024, doi: 10.1016/j.trc.2024.104582.

[14] J. Bao, C. Wu, Y. Lin et al., "A scalable approach to optimize traffic signal control with federated reinforcement learning," *Scientific Reports*, vol. 13, p. 19184, 2023, doi: 10.1038/s41598-023-46074-3.

[15] M. Li, X. Pan, C. Liu et al., "Federated deep reinforcement learning-based urban traffic signal optimal control," *Scientific Reports*, vol. 15, p. 11724, 2025, doi: 10.1038/s41598-025-91966-1.

[16] M. Yazdani, M. Sarvi, S. A. Bagloee, N. Nassir, J. Price, and H. Parineh, "Intelligent vehicle pedestrian light (IVPL): A deep reinforcement learning approach for traffic signal control," *Transportation Research Part C: Emerging Technologies*, vol. 149, p. 103991, 2023, doi: 10.1016/j.trc.2022.103991.

[17] A. Ren, G. Zhang, C. Chang, and D. Huang, "Two-step deep reinforcement learning for traffic signal control to improve pedestrian

safety using connected vehicle data," *Accident Analysis & Prevention*, vol. 222, p. 108161, 2025, doi: 10.1016/j.aap.2025.108161.

[18] D. Garg, M. Chli, and G. Vogiatzis, "Fully-Autonomous, Vision-based Traffic Signal Control: from Simulation to Reality," in *Proc. 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Online, 2022, pp. 1-9.

[19] G. Zhang, X. Zhao, and G. Wang, "End-to-End Reinforcement Learning for Traffic Signal Control: Real-Time Video to Signal Decisions," 2024. [Online]. Available: https://openreview.net/forum?id=eM5dar35Ys

[20] P. Michailidis, I. Michailidis, C. R. Lazaridis, and E. Kosmatopoulos, "Traffic Signal Control via Reinforcement Learning: A Review on Applications and Innovations," *Infrastructures*, vol. 10, p. 114, 2025, doi: 10.3390/infrastructures10050114.

[21] M. Elharoun, S. El-Badawy, E. Shwaly, and U. Shahdah, "Adaptive traffic signal control using deep reinforcement learning: A multi-objective approach for single and multi-intersection scenarios," *IATSS Research*, vol. 49, pp. 481-492, 2025, doi: 10.1016/j.iatssr.2025.10.004.

[22] A. Saadi, N. Abghour, Z. Chiba et al., "A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control," *Journal of Big Data*, vol. 12, p. 84, 2025, doi: 10.1186/s40537-025-01104-x.

[23] D. V. Gadasin and A. V. Shvedov, "Application of the transportation problem for load balancing under fuzzy initial data," *T-Comm*, vol. 18, no. 1, pp. 13-20, 2024, doi: 10.36724/2072-8735-2024-18-1-13-20.

[24] A. V. Shvedov, D. V. Gadasin, and A. V. Alyoshintsev, "Segment routing in data transmission networks," *T-Comm*, vol. 16, no. 5, pp. 56-62, 2022, doi: 10.36724/2072-8735-2022-16-5-56-62.

[25] D. V. Gadasin, A. V. Shvedov, and I. A. Kuzin, "Three-dimensional reconstruction of an object from a single image using deep convolutional neural networks," *T-Comm*, vol. 16, no. 7, pp. 29-35, 2022, doi: 10.36724/2072-8735-2022-16-7-29-35.

АДАПТИВНОЕ УПРАВЛЕНИЕ СВЕТОФОРАМИ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С УЧЕТОМ ПОТОКОВ ТРАНСПОРТНЫХ СРЕДСТВ И ПЕШЕХОДОВ

Полянцева Ксения Андреевна, *Московский технический университет связи и информатики, Москва, Россия,* **\k.a.poliantseva@mtuci.ru**

Городничев Михаил Геннадьевич, *Московский технический университет связи и информатики, Москва, Россия,* **m.g.gorodnichev@mtuci.ru**

Кулагин Владимир Петрович, *Московский технический университет связи и информатики, Москва, Россия,* **v.p.kulagin@mtuci.ru**

Аннотация

В работе рассматривается задача адаптивного управления светофорным объектом с совместным учётом транспортных и пешеходных потоков. Предложен метод, основанный на алгоритме Deep Q-Network, в котором агент принимает решение о сохранении или переключении фазы светофора на основе текущего состояния перекрёстка. Состояние включает загруженность полос движения, количество пешеходов в зоне перехода и текущую фазу регулирования. Основной вклад работы связан с разработкой многокомпонентной адаптивной функции вознаграждения. В неё включены задержки транспортных средств, ожидание пешеходов, пропускная способность перекрёстка и штраф за избыточные переключения фаз. При накоплении пешеходов в зоне перехода пешеходная компонента усиливается, что позволяет агенту избегать автомобиль-центричной стратегии управления и учитывать интересы разных участников движения. Экспериментальная проверка выполнена в среде микроскопического моделирования SUMO. Предложенный метод сравнивался с фиксированным управлением светофором. Полученные результаты показали снижение среднего времени ожидания пешеходов и транспортных средств, сокращение длины очередей и рост пропускной способности перекрёстка. Результаты подтверждают, что обучение с подкреплением в сочетании с адаптивной функцией вознаграждения может быть использовано для более гибкого и сбалансированного управления городским перекрестком.

Ключевые слова: *адаптивное управление светофорами, обучение с подкреплением, Deep Q-Network, транспортные потоки, пешеходные потоки, SUMO, интеллектуальные транспортные системы*

Литература

1. *Polyantseva K. A., Gorodnichev M. G., Moseva M. S., Potapchenko T. D.* On the Applicability of Neural Networks in the Tasks of Detecting Dangerous Movement // Proc. Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russia, 2019, p. 8840654, doi: 10.1109/WECONF.2019.8840654.

2. *Gorodnichev M. G., Polyantseva K. A., Yashina M. V.* Informational Aspects of Automated Monitoring of Dangerous Driving // Proc. Systems of Signals Generating and Processing in the Field of on Board Communications (SOSG), Moscow, Russia, 2019, p. 8706721, doi: 10.1109/SOSG.2019.8706721.

3. *Polyantseva K., Gorodnichev M., Mitrokhin Y., Moseva M.* On the Problem of Developing a Realistic Road Infrastructure Simulator for Reinforced Learning // *Proc. Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russian Federation, 2023, pp. 1-8, doi: 10.1109/WECONF57201.2023.10147983.

4. *Noaeen M., Naik A., Goodman L., Crebo J., Abrar T., Abad Z. S. H., Bazzan A. L. C., Far B.* Reinforcement learning in urban network traffic signal control: A systematic literature review // Expert Systems with Applications, vol. 199, p. 116830, 2022, doi: 10.1016/j.eswa.2022.116830.

5. *Zhao H., Dong C., Cao J., Chen Q.* A survey on deep reinforcement learning approaches for traffic signal control // Engineering Applications of Artificial Intelligence, vol. 133, Part A, p. 108100, 2024, doi: 10.1016/j.engappai.2024.108100.

6. *Bouktif S., Cheniki A., Ouni A., El-Sayed H.* Deep re-inforcement learning for traffic signal control with consistent state and reward design approach // Knowledge-Based Systems, vol. 267, p. 110440, 2023, doi: 10.1016/j.knosys.2023.110440.

7. *Cai C., Wei M.* Adaptive urban traffic signal control based on enhanced deep reinforcement learning // Scientific Reports, vol. 14, p. 14116, 2024, doi: 10.1038/s41598-024-64885-w.

8. *Wang L., Zhang G., Yang Q., Han T.* An adaptive traffic signal control scheme with Proximal Policy Optimization based on deep reinforcement learning for a single intersection // Engineering Applications of Artificial Intelligence, vol. 149, p. 110440, 2025, doi: 10.1016/j.engappai.2025.110440.

9. *Jiang Q., Qin M., Shi S., Sun W., Zheng B.* Multi-agent reinforcement learning for traffic signal control through universal communication method // Proc. 31st International Joint Conference on Artificial Intelligence (IJCAI-ECAI), Vienna, Austria, 2022, pp. 3854-3860.

10. *Song X., Zhou B., Ma D.* Cooperative Traffic Signal Control Through A Counterfactual Multi-Agent Deep Actor Critic Approach // SSRN, 2022, doi: 10.2139/ssrn.4021959.

11. *Yang S.* Hierarchical graph multi-agent reinforcement learning for traffic signal control // Information Sciences, vol. 634, pp. 55-72, 2023, doi: 10.1016/j.ins.2023.03.087.

12. *Bie Y., Ji Y., Ma D.* Multi-agent Deep Reinforcement Learning collaborative Traffic Signal Control method considering intersection heterogeneity // Transportation Research Part C: Emerging Technologies, vol. 164, p. 104663, 2024, doi: 10.1016/j.trc.2024.104663.

13. *Wang T., Zhu Z., Zhang J., Tian J., Zhang W.* A large-scale traffic signal control algorithm based on multi-layer graph deep reinforcement learning // Transportation Research Part C: Emerging Technologies, vol. 162, p. 104582, 2024, doi: 10.1016/j.trc.2024.104582.

14. *Bao J., Wu C., Lin Y.* et al., "A scalable approach to opti-mize traffic signal control with federated reinforcement learning // Scientific Reports, vol. 13, p. 19184, 2023, doi: 10.1038/s41598-023-46074-3.

15. *Li M., Pan X., Liu C.* et al., "Federated deep reinforcement learning-based urban traffic signal optimal control // Scientific Reports, vol. 15, p. 11724, 2025, doi: 10.1038/s41598-025-91966-1.

16. *Yazdani M., Sarvi M., Bagloee S.A., Nassir N., Price J., Parineh H.* Intelligent vehicle pedestrian light (IVPL): A deep reinforcement learning approach for traffic signal control // Transportation Research Part C: Emerging Technologies, vol. 149, p. 103991, 2023, doi: 10.1016/j.trc.2022.103991.

17. *Ren A., Zhang G., Chang C., Huang D.* Two-step deep reinforcement learning for traffic signal control to improve pe-destrian safety using connected vehicle data // Accident Analysis & Prevention, vol. 222, p. 108161, 2025, doi: 10.1016/j.aap.2025.108161.

18. *Garg D., Chli M., Vogiatzis G.* Fully-Autonomous, Vision-based Traffic Signal Control: from Simulation to Reality // Proc. 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Online, 2022, pp. 1-9.

19. *Zhang G., Zhao X., Wang G.* End-to-End Reinforce-ment Learning for Traffic Signal Control: Real-Time Video to Signal Decisions // 2024. [Online]. Available: https://openreview.net/forum?id=eM5dar35Ys

20. *Michailidis P., Michailidis I., Lazaridis C. R., Kosmatopoulos E.* Traffic Signal Control via Reinforcement Learning: A Review on Applications and Innovations // Infrastructures, vol. 10, p. 114, 2025, doi: 10.3390/infrastructures10050114.

21. *Elharoun M., El-Badawy S., Shwaly E., Shahdah U.* Adaptive traffic signal control using deep reinforcement learning: A multi-objective approach for sin-gle and multi-intersection scenarios // IATSS Research, vol. 49, pp. 481-492, 2025, doi: 10.1016/j.iatssr.2025.10.004.

22. *Saadi A., Abghour N., Chiba Z.* et al., "A survey of rein-forcement and deep reinforcement learning for coordination in intelligent traffic light control // Journal of Big Data, vol. 12, p. 84, 2025, doi: 10.1186/s40537-025-01104-x.

23. *Гадасин Д.В., Шведов А.В.* Применение транспортной задачи для балансировки нагрузки в условиях нечеткости исходных данных // T-Comm: Телекоммуникации и транспорт. 2024. -Т. 18, № 1. С. 13-20. DOI 10.36724/2072-8735-2024-18-1-13-20.

24. *Shvedov A.V., Gadasin D.V., Alyoshintsev A.V.* Segment routing in data transmission networks // T-Comm: Телекоммуникации и транспорт, vol. 16, no. 5, pp. 56-62, 2022, doi: 10.36724/2072-8735-2022-16-5-56-62.

25. *Гадасин Д.В., Шведов А.В., Кузин И.А.* Трехмерная реконструкции объекта по одному изображению с использованием глубоких свёрточных нейронных сетей // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16, № 7. С. 29-35. DOI 10.36724/2072-8735-2022-16-7-29-35.

Информация об авторах:

Полянцева Ксения Андреевна, *кандидат технических наук, Московский технический университет связи и информатики, Москва, Россия*

Городничев Михаил Геннадьевич, *кандидат технических наук, доцент, Московский технический университет связи и информатики, Москва, Россия*

Кулагин Владимир Петрович, *доктор технических наук, профессор, Московский технический университет связи и информатики, Москва, Россия*