Explainable Deep Reinforcement Learning for Patrol Speed Control of Rail-Guided Robot System

Hosun Lee[1], Jaesung Kwon[2], Nak Young Chong[1], and Woosung Yang[2]

[1] Japan Advanced Institute of Science and Technology, Ishikawa, Japan,
Hosun.JAIST@gmail.com
[2] Kwangwoon University, Seoul, Republic of Korea,
dreamrize@kw.ac.kr

Abstract. Intelligent facility management systems can operate autonomously, reducing the workload for workers. However, existing machine learning-based systems do not provide users with information that serves as an understandable basis for decisions on tasks performed autonomously, making it difficult for users to trust the system. Facility management systems, which require high reliability in safety management, need functions that explain autonomously performed tasks as autonomy increases to support user understanding. This paper proposes an explainable deep reinforcement learning framework for a rail-guided patrol robot, utilizing the Deep Deterministic Policy Gradient (DDPG) algorithm and the Grad-CAM algorithm. The DDPG-based framework autonomously controls the patrol speed of the robot based on environmental information at the site. The Grad-CAM-based XAI technique is applied to the patrol speed control framework to visually highlight environmental factors that significantly impact the robot's decision-making. Through the proposed safety patrol robot control framework, the robot adjusts its speed in response to changes in the patrol environment, projecting the specific locations that influenced speed changes onto image data at the points where these changes occurred.

Keywords: Rail-guided robot, Safety patrol robot, Explainable AI, Reinforcement learning, Deep Deterministic Policy Gradient algorithm, Patrol speed control

1 Introduction

There are increasing cases of safety management methods using robots to overcome the limited information and utilization of fixed monitoring devices such as fire alarms, air qualifiers, and CCTV. By automating patrol tasks using robots, repetitive workloads can be reduced, and the speed of on-site response can be increased[1, 2]. Robots moving autonomously in facilities can monitor safety conditions and detect dangerous situations to report and deliver alarms. The patrol performance will be determined by the level of autonomy and reliability of system operation from the administrator's point of view. Therefore, the robot determines the patrol plan variably according to the situation of the site, and the decision should be explained to the administrator. In this study, we propose an Explainable Deep Reinforcement Learning framework based on Deep Deterministic Policy Gradient(DDPG) algorithm[3] and Grad-CAM algorithm[4] for a rail-guided

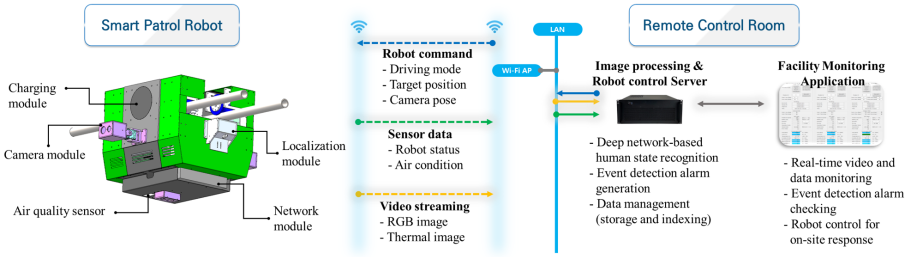

Fig. 1. System overview: configuration of smart patrol robot, the definition of data, and service of remote control room

patrol robot that controls the patrol speed of the robot according to the situation information in the field and provides the visual information that has a high impact on the important control values.

2 Rail-guiede Smart Patrol Robot for Facility Safety Management

The previously developed system for managing the safety of a multi-use facility using a rail-guided patrol robot, which is the subject of this study, is configured as shown in Fig.1 [2].

- **Smart patrol robot**: The patrol robot is designed to drive on a circular pipe rail, therefore, the position of the robot can be controlled along the rail path by driving forward and backward. A pan-tilt actuator has one RGB and one thermal camera to capture the site situation. A set of air qualifiers is also attached to monitor the environmental conditions such as temperature, humidity, CO_2, Volatile Organic Compounds(VOC), and Particulate Matter(PM1.0, PM2.5, PM10).
- **Remote control room**: In the remote control room, the data processing server gathers all images and sensor data from the smart patrol robot. Deep network-based detection services are implemented to detect emergencies by processing the RGB image data. Conditional alarms are defined as cases of high temperatures and abnormality of air conditions by monitoring the thermal image and air quality sensor data.

It is possible to reduce monitoring workload and improve on-site response through the automation service of patrol tasks and alarm generation provided through the developed system operated with the integration of the smart patrol robot and the remote server.

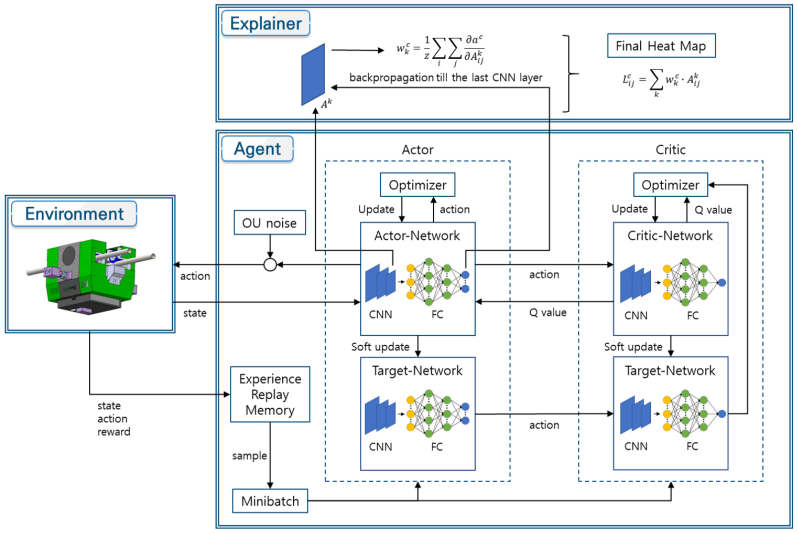

Fig. 2. Patrol speed control model using DDPG algorothm and Grad-CAM algorhtm

3 Explainable Deep Reinforcement Learning framework based on DDPG and Grad-CAM algorithm

3.1 Deep Deterministic Policy Gradient(DDPG) algorithm for the patrol speed control

To effectively monitor the situation of the target facility, we propose a control model based on the Deep Deterministic Policy Gradient (DDPG) algorithm as a method of training the optimum robot patrol speed based on the amount of information on the site. DDPG is an actor-critic algorithm-based reinforcement learning algorithm that applies Deep Neural Network techniques to the DPG (Deterministic Policy Gradient) algorithm. The existing policy-based reinforcement learning algorithm can only handle discrete actions where the policy outputs the probability of taking the possible actions. However, the DPG algorithm deterministically determines and outputs the action value among continuous values without using the probability distribution of the policy to determine the action. Therefore, it is possible to output the action value in the real number range without making a choice. Regarding the problem of determining the robot's patrol speed according to the site's condition, which is the objective of this study, the DDPG algorithm can be used to design a control model that outputs continuous robot control values with image data of the site as input (Fig.2). An experience replay method is used to prevent gradients from being biased due to the temporal correlation of training data. The experience is not directly used for training, but stored in the replay buffer, and N samples are randomly extracted from the buffer. Next, the target actor neural network and the target critical neural network are created. When the Critic Neural Network is updated, the loss function is used, and at this time, the time difference target (TD-Target) is affected and changed. When

the Actor Neural Network is updated, the gradient value of the policy parameter is used. This gradient function is shown in Eq.1.

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_t \left[\nabla_a Q(s,a;\theta^Q)|_{s=s_t, a=\mu(s_t|\theta^\mu)} \nabla_{\theta^\mu} \mu(s;\theta^\mu)|_{s=s_t} \right] \tag{1}$$

θ^Q notes the parameters of the Critic Network and θ^μ notes the parameters of the Actor Network. This makes it impossible to learn stably because the target is constantly changing. Therefore, put the target actor neural network and the target critical neural network separately so that they slowly follow the parameters of this neural network(Soft target update). The original behavior is random, so the given environment must be properly explored, but DDPG uses a deterministic policy, so the policy is uniform. Orstein-Uhlenbeck(OU) noise is used to solve the uniform problem. Optimally controlling the patrol speed of the robot using the amount of information from the image data of the site means that the robot is controlled according to the detecting performance to secure the optimal processing time for each local site. The amount of information in each place can be calculated as the entropy of the transferred image, therefore, the model can be trained to change the patrol speed by defining the reward as an equation of the entropy value (Eq.2).

$$reward = \left| E_d - E_f \left(\frac{v}{fps \times w_{FOV}} \right) \right| \tag{2}$$

The amount of obtained information is defined as the entropy equal to the ratio of the newly added area to the entropy of the current frame image. The ratio of the newly added area is the ratio of the traveled distance between frames, w_{FOV}, out of the covered distance with one frame, v/fps. Therefore, the reward function is designed as the difference between the amount of obtained information to the amount of desired information, E_d.

3.2 Grad-CAM algorithm for a visual explanation

To better explain the operation of the trained model to the administrator, a visual explanation is generated by applying an eXplainable AI (XAI) method, the Grad-CAM algorithm. The proposed DDPG-based model includes a CNN to process image data, but since the spatial information of convolutional features is lost in the FC layer, it is difficult to describe information that is important for robot commands in the input image. As shown in Fig.2, the explainer uses the gradient information of the last convolutional layer of CNN to assign importance to each neuron for the particular actions of the DDPG model. $A^k \in \mathbb{R}^{u \times v}$ denotes the feature map of the kth channel of the last CNN layer. w_k^c means partial linearization of the deep network and represents the importance of the feature map of k channels for a particular action, c. Where, $\frac{\partial a^c}{\partial A_{ij}^k}$ can be calculated via backpropagation.

4 Evaluation

In order to train and evaluate the proposed model, a warehouse virtual environment managed by the robot is created based on Unity as shown in Fig.3. Unity warehouse and UMA packages are used to configure an environment similar to

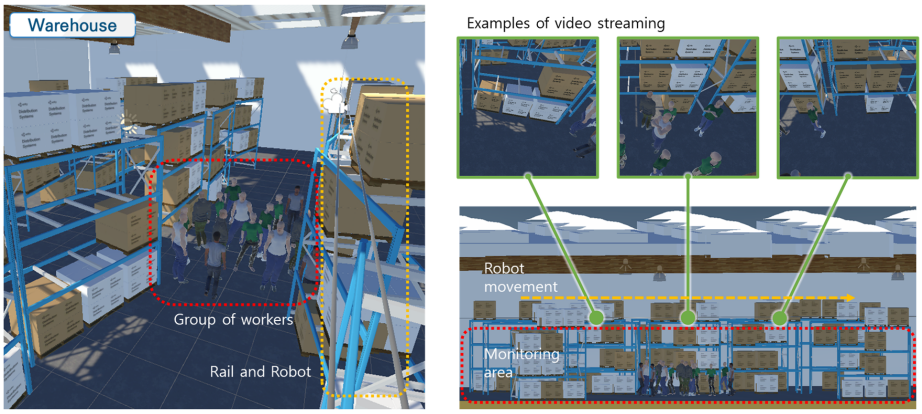

Fig. 3. Simulation environment.

the actual market. The aisle area is divided into an area where a group of workers or visitors is concentrated and an area without it. The simulation environment is constructed using the extensional packages: ML-Aagents, Warehouse, and Unity Multipose Avatar(UMA)[5–7]. The ML-Agents package supports environments for building and training intelligent agents. It provides Python API to implement the deep learning pipeline based on PyTorch. The specification of the computer is set as Intel i7-8700K@3.70GHz, RTX 3080Ti GPU, 32GB RAM. Fig.4 shows the simulation result of the trained controller. The trained controller takes the image captured by the robot from the environment as input. The patrol speed of the robot was controlled according to the output value from the trained controller. The speed comparison result at 7 locations in the monitoring area with sample images shows that the robot patrols at the highest speed in the empty area and moves at the lowest speed in the crowded area with workers and visitors. Calculating the entropy of each sample image in the upper row shows that the amount of information in the environment is represented. The images in the bottom row are Grad-CAM result images. It expresses the degree of influence in determining the patrol speed of the robot on the input image. It shows the influence is significant in the 2nd and 3rd sections where the speed continues to change.

5 Conclusion

This research proposes a method to increase the autonomy of an automated system to reduce the workload of workers managing the safety of facilities and to generate information to increase reliability with users when autonomy is increased. Deep Deterministic Policy Gradient (DDPG) algorithm is applied to operate the patrol robot with the optimal control of the patrol speed. The visually highlighted environmental factors that significantly impact the robot's decision-making are provided by applying the Grad-CAM-based XAI technique to the patrol speed control framework. The designed model can be trained by defining the reward function based on the entropy to maintain the obtained information. The performance of the trained controller shows that the patrol speed is efficiently

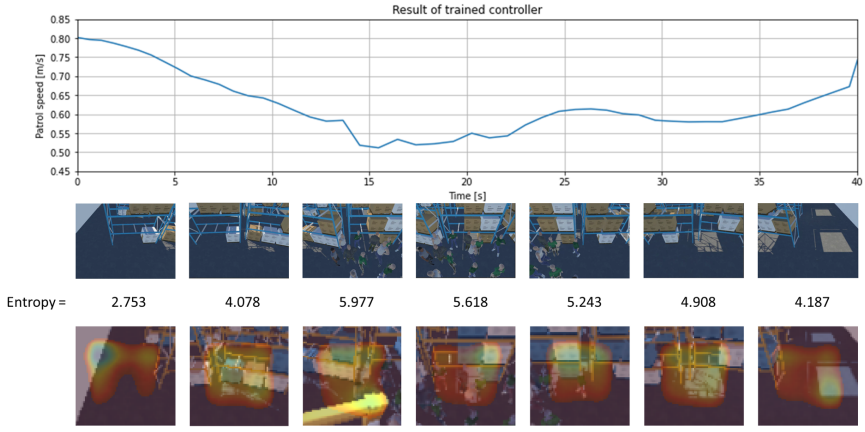

Fig. 4. Result of simulation.

controlled according to the environment and the the visual information about the decision-making is provided. In future work, extensive experiments and analysis will be performed to evaluate the performance of the proposed framework. This developed model can be further improved by employing additional robot motions for facility management functions.

Acknowledgement

This work was partly supported by grants funded by the National Research Foundation of Korea (NRF-2022M3C1A3099340).

References

1. Halder S, Afsari K.: Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review. Applied Sciences, 13(4):2304, 2023
2. Lee H., Kwon J., Shin M., Lee S., Chong N. Y., Yang W.: Development of Rail-guided Smart Patrol System for Surveillance and Monitoring of Facilities Safety. 2023 IEEE/SICE International Symposium on System Integration (SII), 1–6, 2023
3. Lillicrap T. P., Hunt J. J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D.: Continuous control with deep reinforcement learning. arxiv preprint arxiv:1509.02971 (2015)
4. Selvaraju R. R. and Cogswell M. and Das A., Vedantam R, Parikh D., Batra D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 618–626, 2017
5. Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., Lange, D.: Unity: A general platform for intelligent agents. arXiv preprint arXiv:1809.02627 (2020)

6. Unity-Technology: Robotics-Warehouse https://github.com/Unity-Technologies/Robotics-Warehouse
7. UMA2 - Unity Multipurpose Avatar https://assetstore.unity.com/packages/3d/characters/uma-2-unity-multipurpose-avatar-35611