論文の要旨

題目　Optimizing Controllers of Swarm Robotic Systems with Deep Reinforcement Learning
（深層強化学習によるスワームロボットのコントローラの最適化）

氏名　NIE XIAOTONG

In recent years, Deep Reinforcment Learning (DRL) has demonstrated great potential in designing controllers to various static environment tasks, such as playing video games. However, it is difficult for traditional DRL to learn effective policies in dynamic environments due to the lack of complete observability and the non-stationarity of the environment. Swarm Robotic System (SRS) is the study of a large group of autonomous robots that operate without relying on a centralized controller and global information. In SRSs, each robot only has a local observation of the environment and other robots are treated as part of the environment. This characteristic results in a highly dynamic environment that poses significant challenges for DRL algorithms. Therefore, this thesis focuses on optimizing controllers of SRS with DRL.

To address this challenge, one of the efficient ways is to integrate with other algorithms, such as curriculum learning, to improve the performance of control algorithms. On the other perspective, enhancing the understanding of DRL through explainable algorithms is also critical. By understanding decision-making processes and policy characteristics, potential issues will be discovered and improved. In this thesis, three contributions are presented to the field of SRS and DRL.

Firstly, when faced with complex tasks, DRL is insufficient for directly training end-to-end controllers. To address this issue, curriculum learning has emerged as a promising solution. However, a traditional manually designed curriculum limits the pace of training and heavily relies on the designer's experience. Therefore, a novel automatic curriculum learning method called Self-Teaching Automatic Curriculum Learning (STACL) is proposed. In order to illustrate the effectiveness of STACL, this study uses a collective wall-jumping task, in which the robots have to jump over the high wall collectively and reach the goal as quickly as possible. The proposed algorithm integrates robot training with curriculum scheduling in one neural network. The reward function can calculate the learning rate for different curricula, then select the next subtask to be trained automatically for the next episode. The proposed method can ensure that the neural network remains in an optimal state for learning. The proposed approach is compared with the manual CL, random CL, and a conventional RL algorithm. Simulation resluts demonstrate that the proposed method has the quickest convergence speed since it can automatically schedule lessons and is unaffected by manual settings. Additionally, we also performed experiments to examine the flexibility of the proposed approach.

Secondly, this thesis presents how to utilize DRL to address a decision-making problem in a multi-autonomous vehicle task. In this task, multi-autonomous vehicles will be formulized as a SRS controlled by DRL algorithm. Other traffic participants, i.e, environmental vehicles, are seen as part of the environment. The positions and actions of environmental vehicles are unpredictable, and their movements may affect the decisions of autonomous vehicles controlled by DRL, which makes the environment more dynamic and dangerous. Therefore, it is necessary

to equip autonomous vehicles with a security assurance mechanism. The proposed method utilizes time-to-collision (TTC) as the feature representation and proposes a TTC-based safety check system. In this study, a ramp merging task is used to illustrate the effect. The action output by the DRL controller would be replaced with a safer action chosen by the safety check system when an agent detects a potential collision. Simulation results show that the proposed method can effectively improve the arrival rate and reduce the collision rate, even in the case of dense traffic situations. Furthermore, we also examine the performance of the safety check system with different time thresholds.

Thirdly, this thesis develop an explainable reinforcement learning approach. The lack of interpretability problem limits the understanding and optimization of the model's decision-making in dynamic environments. In this study, a deep Q-learning algorithm is applied to develop controllers for a SRS that take raw camera images as input. Three experiments are conducted to visualize the policies learned by deep Q-network. The first experiment proposes a network structure with several deconvolutional layers to view the neural network's feature map during various training stages. At each stage of the learning process, this approach promotes a more comprehensive understanding of the underlying control strategy. The second experiment employs a saliency map method Grad-CAM to determine which state variables robots attend to view. Lastly, the third experiment utilizes a perturbation-based visualization method to evaluate the fault tolerance of the controller. Simulation results show that the proposed approach can interpret the control policies.

Overall, the proposed methods can optimizing controllers of SRS with DRL, provide promising solutions for addressing complex tasks in dynamic environments that traditional DRL approaches struggle with.

This thesis consists of six chapters. The following is a summary of each chapter.

Chapter 1 introduces the thesis by providing background information on the research.

Chapter 2 first gives a review of SRS. Then, DRL and several related optimazation techniques are presented.

Chapter 3 proposed a novel automatic curriculum learning method called Self-Teaching Automatic Curriculum Learning to solve the issue that DRL is insufficient for directly training end-to-end controllers.

Chapter 4 presented how DRL is utilized to address a decision-making problem in a multi-autonomous vehicle task, where multi-autonomous vehicles are formulized as a SRS controlled by DRL algorithm.

Chapter 5 utilizes several visualization approaches to improve the understanding of the control strategy.

Chapter 6 concludes the thesis and discusses future research directions.