

강화학습을 이용한 실내 자율주행 시스템

이기성¹, 이종찬^{2*}

¹호원대학교 컴퓨터게임학과, ²군산대학교 소프트웨어학부

Indoor autonomous driving system using reinforcement learning

Gi-Sung Lee¹, Jong-Chan Lee^{2*}

¹Department of Computer & Game, Howon University

²School of Software, Kunsan National University

요약 자율주행은 미래 이동체의 핵심 기술 중 하나로 강화학습은 주행 에이전트가 환경과 상호작용하며 최적의 행동을 학습하는 데 유용한 도구이다. 본 논문은 실내 환경에서 강화학습을 이용하여 자율주행을 구현하는 방법으로, 실시간으로 변화하는 실내 자율주행 환경을 인식하고 주변 사물의 출현에 대응하여 회피하며 사전 설정된 다수의 목적지에 도달하기 위한 방법을 제안한다. 이 기술은 라이다 센서와 영상 센서를 탑재한 이동체가 실내를 반복적으로 순회하면서 강화학습 모듈이 지형과 지물을 학습함으로써 실내 환경을 스스로 인지한다. 우선, 주행 환경을 2D 시뮬레이션으로 모델링하고, 주행 경로와 장애물을 정확하게 표현하며, 강화학습 주행 에이전트는 강화학습 알고리즘을 사용하여 환경에서 행동을 수행하고 보상을 받는다. 에이전트는 자율주행을 위한 최적 정책을 학습하고, 에피소드를 반복하면서 환경에서 행동을 수행하고 보상을 최대화하기 위한 정책을 학습한다. 실제 강화학습 모듈을 탑재한 자율주행 차량은 실내 주행을 수행하며 환경의 변화에 적응한다. 연구 결과, 강화학습을 통해 자율주행 시스템이 동적 환경에서 안정적이고 효율적으로 작동할 수 있음을 입증하였다. 이러한 연구는 자율주행 기술의 발전과 미래 도시 모빌리티에 기여할 것으로 기대된다.

Abstract Autonomous driving is one of the key technologies for future mobile vehicles, and reinforcement learning is a useful tool for driving agents interacting with the environment and learning optimal behavior. This paper presents a method of applying reinforcement learning to autonomous driving in an indoor environment. It recognizes an indoor autonomous driving environment that changes in real time; it deals with the unexpected appearances of objects, and returns to a number of preset destinations. We suggest a method to reach. This technology uses reinforcement learning in which a mobile device, equipped with a lidar sensor and an image sensor, is programmed to recognize the indoor environment on its own by repeatedly traveling around the room to learn the layout and features. First, the driving environment is modeled in a 2D simulation; the driving path and obstacles are accurately expressed, and the reinforcement learning driving agent uses a reinforcement learning algorithm to perform actions in the environment and get rewards. The agent learns the optimal policy for autonomous driving, performs actions in the environment through repeated episodes, and learns a policy to maximize rewards. The final model is deployed in an actual autonomous vehicle to drive indoors and adapt to changes in the environment. Results demonstrate that this autonomous driving system can operate stably and efficiently in a dynamic environment through reinforcement learning. This research is expected to contribute to the development of autonomous driving technology and future urban mobility.

Keywords : Autonomous Driving, Lidar, Reinforcement Learning, Mobile Vehicles, Image Sensor

본 논문은 2023년도 호원대학교 지원에 의하여 출판되었음.

*Corresponding Author: Jong-Chan Lee(Kunsan National Univ.)

email: chan2000@kunsan.ac.kr

Received September 1, 2023

Revised October 5, 2023

Accepted October 6, 2023

Published October 31, 2023

1. 서론

현대 사회에서 자율주행 기술은 무엇보다도 빠르게 발전하고 있는 분야 중 하나로, 자율주행 차량은 운전의 부담을 덜어주고 도로 안전성을 향상하며 교통 효율성을 증가시킨다. 특히, 실내 환경에서의 자율주행은 미래 스마트 도시 및 지능형 건물의 중요한 부분이 될 것으로 예상된다. 이러한 자율주행 시스템은 다양한 응용 분야에서 혁신적인 변화를 가져올 것으로 기대되는데, 그중에서도 강화학습은 자율주행 분야에서 주목받는 중요한 연구 중 하나이다[1,2].

강화학습은 기계학습의 한 분야로, 에이전트가 환경과 상호 작용하며 보상을 최대화하도록 학습하는 방식이다. 자율주행 차량의 경우, 에이전트는 차량이 주행하면서 다양한 의사결정을 해야 하며, 이러한 의사결정은 주행 경로 선택, 속도 조절, 교통 상황 대응 등 다양한 측면에서 이루어진다. 이러한 복잡한 환경에서 최적의 의사결정을 하기 위해서는 심층 학습 알고리즘 중 하나인 강화 학습이 효과적으로 활용될 수 있다. 강화학습은 다른 기계학습 방법과 비교하여, 획득한 데이터에 라벨을 붙일 필요가 없으므로 학습 과정이 크게 축진되고 오프라인 원칙을 통해 훈련 에이전트는 경험을 통해 학습할 수 있으며, 다소 복잡한 프로세스를 통하여 중요한 개선 효과를 얻을 수 있다[3-5].

일반적인 자율주행 아키텍처의 제어 계층은 차량 제어 및 내비게이션 기능을 구현하는 프로세스로 구성된다[6]. 제어 계층은 차량 안정성을 보장하고 주변 상황, 업무 내용, 실내 운행 시나리오 등 다양한 환경 상황에 적합한 매커니즘을 지원한다. 지도를 바탕으로 경로를 계획하고 이동체의 위치를 매핑하는데, 인공지능은 자율주행 아키텍처를 통하여 이동체 감지, 다중 객체 추적(MOT) 및 환경 예측과 같은 다양한 프로세스를 처리하고 자율주행차의 현재 상황을 평가하여 가장 안전한 결정을 수행한다[7,8]. 자율주행을 위한 DRL(Deep Reinforced Learning) 기반 알고리즘은 Markov Decision Process(MDP) 기반으로 위의 문제를 해결하는 데 사용되었다. 보상 함수를 최대화하는 것을 목표로 에이전트의 행동을 선택하는 최적의 정책을 계산하여 매우 성공적인 결과를 얻었다[9]. 그리고 컴퓨터 게임 해결[10]이나 간단한 의사결정 시스템[11]과 같은 분야에서 자율주행 측면으로 접근하여 의사결정 문제를 정밀하게 해결하려 하였다.

본 논문에서는 실내 자율주행 분야에서 강화학습을 활용하는 방법을 제안한다. 실내 환경에서 이동체가 강화

학습 기반으로 라이다 센서와 영상 센서 정보를 이용하여, 맵(Map)을 작성하고 자율 학습을 수행하며, 이를 바탕으로 동적 장애물에 대한 회피와 목적지를 탐색하고 도달하는 방법을 제안한다.

2장에서는 제안 시스템의 구조를 기술하고 3장에서는 자율주행 시스템의 세부 내용을 기술한다. 4장에서는 성능을 평가하고, 5장에서는 결론과 추후 연구를 서술한다.

2. 제안 시스템 구조

Fig. 1과 같이, 제안된 실내 자율주행 플랫폼은 이동체의 자율주행을 담당하는 RL(Reinforcement Learning) part와 RL part로부터의 제어 명령으로 구동되는 Driving part로 구성한다. RL part는 자율주행에 사용될 데이터의 집합체인 IAV(Indoor Autonomous Vehicle) Data Area, 실제 자율주행 학습 및 운행 제어를 담당하는 IAV Agent Area, 자율주행 학습 시에 상태 정보 및 보상을 제공하는 IAV Environment Area로 구성된다. 그리고 Driving part는 환경 정보 수집 및 구동 요소로 구성된 IAV Platform Area와 자율주행 플랫폼을 상태 모니터링 및 관리를 위한 IAV Management Area로 구성된다.

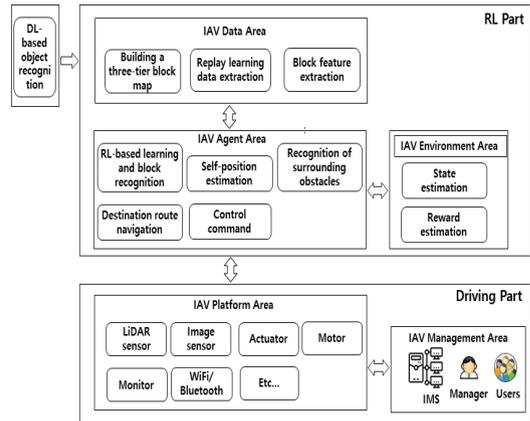


Fig. 1. Service platform for autonomous driving

3. 강화학습 기반의 자율주행 시스템

3.1 블록 특징 데이터 추출

IAV Data Area의 데이터 집합에는 3계층 블록 맵, Replay 학습 데이터, 블록 특징 데이터가 포함된다. 블록

특징 데이터는 라이다 센서(lidar sensor)와 영상 센서(image sensor)에서 생성된 이미지로부터 블록의 특징을 추출한 맵이다. 이런 특징 맵(feature map)을 추출하기 위하여 라이다 센서 이미지의 특징과 영상 센서 이미지의 특징을 이용하여 계층 좌표 및 계층 이미지를 생성한다.

Fig. 2와 같이 각각 영상 센서 이미지와 라이다 2D 스캔 이미지를 CNN(Convolutional Neural Network)의 입력 데이터로, 영상 센서 이미지 특징 맵(픽셀 값)과 라이다 이미지 특징 맵(픽셀 값)을 생성하고, 두 특징 맵을 조합하여 블록 특징 값을 생성한다.

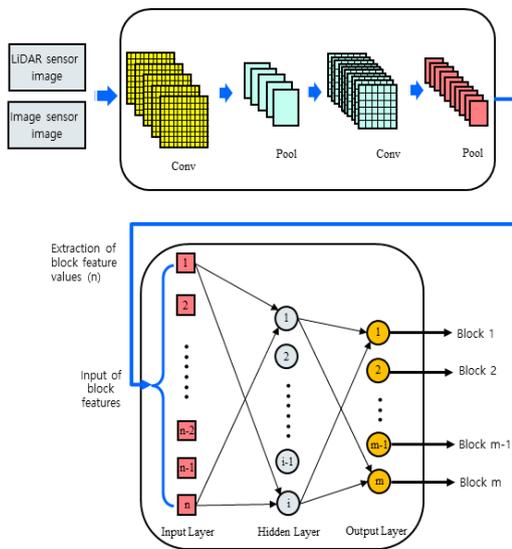


Fig. 2. Extraction of feature maps

3.2 3계층 블록 매칭

3계층 블록 맵 데이터는 가상 맵(virtual map)과 실제 사물을 매핑하기 위하여 생성한다. Fig. 3과 같이, 이동체가 운행되는 실내는 3단 계층 블록으로 구성된다. 영상 블록은 영상 센서로부터 생성된 이미지를 이용하여 실내의 고정된 주변 사물(방, 화장실, 기타)의 위치와 실내의 유동적인 주변 사물(이동 물체 등)의 위치를 나타낸다. 물리 블록(physical block)은 라이다 센서로부터 생성된 이미지를 이용하여 SLAM (Simultaneous Localization and Mapping)에 의하여 생성된 좌표와 실내의 이동 가능 통로, 그리고 실내의 벽 등을 포함한 장애물을 구분하기 위하여 사용된다. 논리 블록(logical block)은 실내 공간의 절대 위치를 나타낸다. Fig. 3과 같이 3단 계층 블록은 논리 블록에서는 가상좌표가 매핑되고, 물리 블록에서는 가상좌표에 벽, 통로, 물체 등이 매핑되며, 영

상 블록에서는 물리 블록에 사무실, 화장실, 기타 정확한 사물 명칭을 매핑한다.

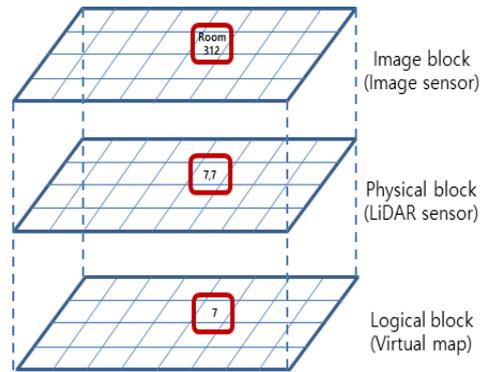


Fig. 3. 3-tier block

IAV Agent Area는 IAV Data Area의 데이터 집합을 이용하여 경로 학습 및 목적지 탐색을 수행한다. 주요 기능은 학습 및 블록 인식, 자기 위치 추정, 주변 장애물 인식, 목적지 경로 탐색 등이 있다. Fig. 4와 같이 블록 특징점을 추출하고 3계층 블록 맵을 구성한 후에, 이를 이용하여 블록 추정 학습 및 목적지 경로 학습을 수행한다.

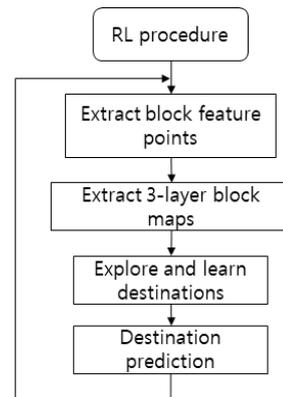


Fig. 4. RL procedure

3.3 목적지 탐색 및 학습

Fig. 5는 목적지 탐색 및 학습을 위한 두 신경망 구조를 보인다. 정책신경망(policy networks)은 정책에 근사하여 상태에 따라 바로 행동(직진, 후진, 정지, 좌회전, 우회전 등)을 선택한다. 상태의 특징 벡터가 입력되고 출력은 각 행동 확률이 된다. 가치 신경망(value networks)은 상태에 따라 행동 가치에 근사한다. 입력으로 상태의 특징 벡터가 입력되고, 각 행동에 대한 가치를 포함하는

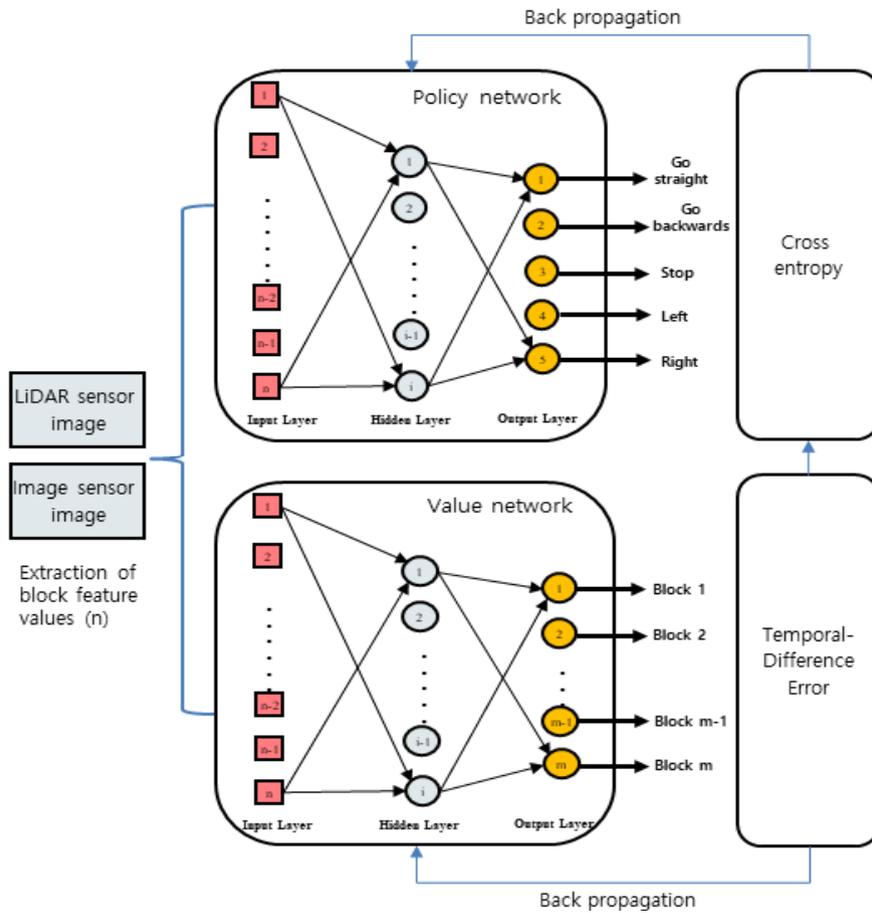


Fig. 5. Policy networks and value networks

큐 함수(위치할 블록의 확률)를 출력한다.

가치 신경망은 시간차 에러(temporal difference error)를 이용하여 신경망을 업데이트한다. 정책신경망은 교차 엔트로피 오차(Cross Entropy Error, CEE)와 시간차 에러의 곱으로 새로운 오류함수를 정의하고 이 오류함수로 정책신경망을 갱신한다.

3.4 이동체 운행 과정

Fig. 6은 이동체의 운행 과정을 보인다. 내비게이션 기능으로서, 라이더의 특징 맵과 영상 센서의 이미지 이용하여 위치를 인식하고 맵을 작성하여 이를 기반으로 목적지까지의 경로를 설정한다. 목적지로 이동 중에 경로를 확인하기 위하여 위치 확인을 수행하고 장애물 확인 시에 경로를 재설정한다. 입력 데이터는 라이더 센서 이미지 데이터, 영상 센서 이미지 데이터이고, 출력은 직진, 후진, 정지, 좌회전, 우회전이다.

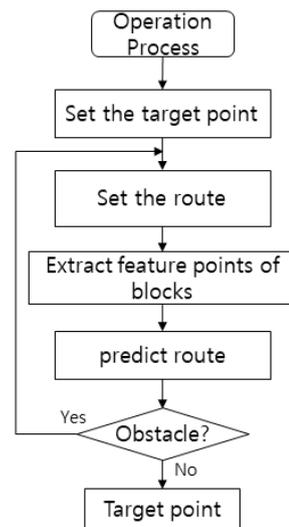


Fig. 6. Operation process

4. 성능 분석

학습이 완료된 모델의 성능을 측정하기 위하여 Fig. 7 과 같은 실험환경에서, 주행 동안의 경로 판단과 장애물 회피 여부를 측정한다. 실험환경은 학교 내 실내 복도 환경이며 장애물을 복도에 배치하여 출발지로부터 목적지까지의 이동체의 운행 상황을 측정한다.

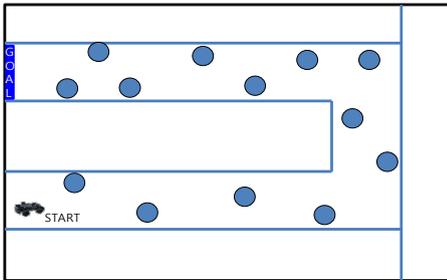


Fig. 7. Obstacle arrangement for Experiment

주행 에이전트는 시뮬레이션 상의 복도를 따라 출발 지점에서 목적지까지 이동하도록 학습한다. 다양한 장애물과 환경 조건을 시뮬레이션하여 현실적인 주행 상황을 재현한다. Fig 8-9와 같이 주행 에이전트의 학습 결과, 분류 정확도는 93.36%, 손실률은 10.1%로서 훈련이 진행될수록 정확도는 증가하고 손실률이 감소함을 알 수 있다. 성공 시의 데이터 셋을 저장하여 검증을 수행하였다. 검증 셋에서의 정확도는 97.06%, 손실률은 7.5%를 나타낸다. 시뮬레이션 환경에서 안정적이고 효과적인 주행을 수행함을 알 수 있다. 주행 에이전트는 장애물을 피하고 목적지까지 안전하게 도달했고, 다양한 환경 조건에 적절하게 대응했으며, 주행 경로의 최적화를 효과적으로 수행했다.

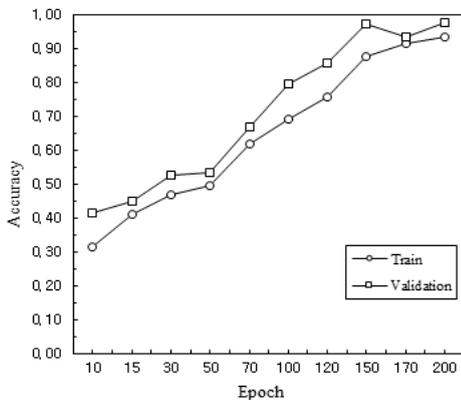


Fig. 8. Model accuracy rate

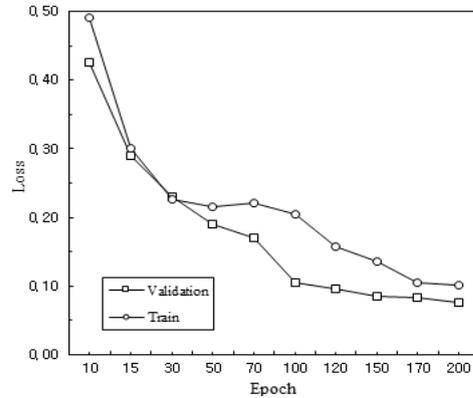


Fig. 9. Model loss rate

5. 결론

강화학습을 이용한 실내 자율주행 시스템은 효율성, 환경 보호, 사회 포용성 및 경제적 이익 등 다양한 측면에서 사회적 공헌을 할 수 있는 기술로 인식되고 있다. 특히 라이다(LiDAR) 센서, 영상 센서가 생성한 이미지는 자율주행 이외에 로봇 기술, 건설 및 산업 자동화, 농업 자동화, 산업 설비 관리, 도시 계획 및 스마트 도시 등으로 활용 영역이 넓어지고 있다.

본 연구에서는 실시간으로 변화하는 실내 자율주행 환경을 인식하고 주변 사물의 출현에 대응하여 회피하며 사전 설정된 다수의 목적지에 도달하기 위한 방법을 제안하였다. 검증 셋에서의 정확도는 97.06%, 손실률은 7.5%로서 시뮬레이션 환경에서 안정적이고 효과적인 주행을 수행함을 알 수 있었다.

향후 연구에서는 학습속도 개선 및 DQN 등의 강화학습 환경에서 추가 제어 연구를 진행할 예정으로 더욱 효과적인 결과를 도출할 것으로 기대된다.

References

- [1] Park Y. S., "A Study on Mobility Furniture Composition for a Segment Model Proposal for Autonomous Driving Cars", Journal of the Korea Academia-Industrial cooperation Society, Vol.23, No.4 pp.370-376, 2022. DOI: <https://doi.org/10.5762/KAIS.2022.23.4.370>
- [2] Kang B. J. and Kim J. W., "Decision Support System of Obstacle Avoidance for Mobile Vehicles", Journal of the Korea Academia-Industrial cooperation Society,

Vol. 19, No.6, pp.639-645, 2018.

DOI: <https://doi.org/10.5762/KAIS.2018.19.6.639>

- [3] Badr B. E., Nabil B., Abdelhakim S. H., et al., "A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving", Journal of King Saud University - Computer and Information Sciences, vol.34, no.9, pp. 7366-7390, 2022.
DOI: <https://doi.org/10.1016/j.jksuci.2022.03.013>
- [4] Estephanos J, Kidus O, Kidus G, et al., "Analysis of Reinforcement Learning in Autonomous Vehicles", pp. 26-29, 2022.
DOI: <https://doi.org/10.1109/CCWC54503.2022.9720883>
- [5] Dimitris M. V., Eleni I. V, John G., "A reinforcement learning model for personalized driving policies identification", Vol.9, No.4, pp. 299-308, 2020.
DOI: <https://doi.org/10.1016/i.ijst.2020.03.002>
- [6] Kendall A, Hawke J, Janz D, et al., "Learning to drive in a day", IEEE International Conference on Robotics and Automation, pp 8248-8254, 2108.
DOI: <https://doi.org/10.48550/arXiv.1807.00412>
- [7] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, et al., "Continuous control with deep reinforcement learning", arXiv preprint arXiv:1509.02971, 2015.
DOI: <https://doi.org/10.48550/arXiv.1509.02971>
- [8] Arango JF, Bergasa LM, Revenga PA, et al., "Drive-by-wire development process based on ros for an autonomous electric vehicle", Sensors 20, 2021.
DOI: <https://doi.org/10.3390/s20216121>
- [9] Mao H, Alizadeh M, Menache I, Kandula S, "Resource management with deep reinforcement learning", In Proceedings of the 15th ACM Workshop on Hot Topics in Networks. pp 50-56, 2016.
DOI: <https://doi.org/10.1145/3005745.3005750>
- [10] Silver D, Hubert T, Schrittwieser J, Antonoglou I, et al., "Mastering chess and shogi by self-play with a general reinforcement learning algorithm", arXiv preprint arXiv:1712.01815, 2017.
DOI: <https://doi.org/10.48550/arXiv.1712.01815>
- [11] Mnih V, Kavukcuoglu K, Silver D, et al., "Human-level control through deep reinforcement learning", Nature 518, pp.529-533, 2015.
DOI: <https://doi.org/10.1038/nature14236>

이 기 성(Gi-Sung Lee)

[중신회원]



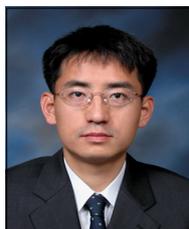
- 1996년 2월 : 숭실대학교 컴퓨터 과학과 (공학석사)
- 2001년 2월 : 숭실대학교 컴퓨터 과학과 (공학박사)
- 2001년 3월 ~ 현재 : 호원대학교 컴퓨터게임학과 교수

<관심분야>

모바일 멀티미디어, 네트워크 보안, 머신러닝

이 중 찬(Jong-Chan Lee)

[정회원]



- 1996년 8월 : 숭실대학교 컴퓨터 과학과 (공학석사)
- 2000년 8월 : 숭실대학교 컴퓨터 과학과 (공학박사)
- 2000년 10월 ~ 2005년 2월 : 한국전자통신연구원 선임연구원
- 2005년 3월 ~ 현재 : 군산대학교 컴퓨터정보공학과 교수

<관심분야>

머신러닝, 빅데이터, 블록체인