

# 자율사물을 위한 심층학습 인공지능 기술 적용 동향

## Application Trends of Deep Learning Artificial Intelligence in Autonomous Things

조준면 (J.M. Cho, jmcho@etri.re.kr)

자율형IoT연구실 책임연구원

### ABSTRACT

Recently, autonomous things, which are pieces of equipment or devices that grasp the context of circumstances on their own and perform actions appropriate for the situation in the surrounding environment, are attracting much research interest. This is because autonomous things are expected to be able to interact with humans more naturally, supersede humans in many tasks, and further solve problems by themselves by collaborating with each other without human intervention. This prospect leans heavily on AI as deep learning has delivered astonishing breakthroughs recently and broadened its range of applications. This paper surveys application trends in deep learning-based AI techniques for autonomous things, especially autonomous driving vehicles, because they present a wide range of problems involving perception, decision, and actions that are very common in other autonomous things.

**KEYWORDS** 인공지능, 심층학습, 딥러닝, 강화학습, 자율주행

## 1. 서론

최근 IoT 분야에서는 자율주행차, 자율비행드론, 자율이동로봇 등 자율사물이 주요 관심사가 되고 있다. 자율성을 가진 기기나 장치들을 통칭하는 자율사물(Autonomous Things)이라는 용어는 미국의 IT 분야 시장조사 및 컨설팅 기관인 가트너가 2019년 10대 전략기술 중 하나로 선정한 이후 본격적으로 언급되고 있다[1].

자율사물의 자율성은 기계가 스스로 주변 상황

과 맥락을 파악해 가장 적합한 기능을 스스로 수행하는 것 정도로 정의할 수 있는데, 다수의 연구 기관과 전문가들은 자율사물 기술이 기존 경직된 프로그래밍 모델이 제공하던 자동화 및 지능화를 넘어 큰 변화를 가져올 수 있다고 예측한다. 즉, 자율사물이 인간의 업무를 대신 담당하거나 인간과 자연스럽게 협력하고 나아가 자율사물 간 협동을 통해 인간의 개입 없이 문제를 해결하는 미래 환경이 가능하다고 진단한다.

자율사물은 비록 공식적으로 통용되는 기준은

\* DOI: <https://doi.org/10.22648/ETRI.2020.J.350601>

\* 본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음[20ZR1100, 산업 및 사회문제 해결을 위한 사물 분산지능 핵심원천 기술 개발].



없지만 가트너가 제시한 자율사물 평가 프레임워크[2]를 통해 유형 및 용도와 자율성 수준을 가늠할 수 있다. 해당 프레임워크에 따르면 자율사물은 로봇, 차량, 드론, 가전, 에이전트 등의 유형으로 분류되며, 용도는 육지, 해상, 공중, 디지털 환경을 기준으로 구분된다. 그리고 자율화 수준은 능력(Capability), 협업(Coordination), 지능(Intelligence) 측면에서 각각 4~5단계로 구분된다.

이러한 자율사물 기술의 원천은 인공지능 기술이다. 자율사물은 사람이 수행하던 업무를 숙련공처럼 처리할 수 있어야 하고, 이를 위해서는 사람처럼 상황을 인지(Perception)하고 판단(Decision)하여 행동(Action)해야 하기 때문이다.

표 1과 같이 산업계와 학계에 걸쳐 인공지능에 대한 다양한 정의가 존재하며[3,8,9], 이러한 정의들은 다양한 관점과 조금씩 다르지만 유용한 측면을 제시한다. 그러나 이들 정의만으로는 실제 인공지능 기술이 적용되어 문제를 풀어내는 내부 메커니즘에 대해 그다지 많은 것을 설명하지 못한다.

본 고에서는 인공지능 기술 특히, 심층학습, 강화학습과 같은 최신 인공지능 기술이 자율사물 개발에 어떻게 적용되고 있는지 그 현황을 살펴본다. 이를 위해 자율주행차를 자율사물의 한 예, 즉 기술적용 분석을 위한 하나의 응용 도메인으로 선정하였는데, 자율주행이 감지(Sensing), 인지, 판단, 계획(Planning), 제어(Control) 등 넓은 영역에 걸쳐 다양한 문제를 제시하기 때문에 자율주행차 도메인에서의 분석 내용이 다른 많은 자율사물에도 동일하게 적용될 수 있기 때문이다.

본 고의 II장에서는 자율주행 시스템이 갖추어야 하는 기능, 즉 자율주행 시스템의 프로세싱 단계와 각 단계에서 해결해야 하는 문제에 대해 정리한다. III장에서는 최신 인공지능 기술을 중심으로 세부 기술들을 분류하고 각 세부 기술의 특성과 문

표 1 인공지능에 대한 다양한 정의 예[3]

정의	연구자/기관
지능적인 기계를 만드는 과학 및 엔지니어링	McCarthy(1956)
인간의 사고와 의사결정, 문제해결, 학습과 같은 활동의 자동화	R. Bellman(1978)
컴퓨터가 사고하도록 만드는 흥분되는 새로운 시도 ... 마음을 가진 기계	J. Haugeland(1985)
계산 모델 활용을 통한 정신적 역량 연구	E. Charniak, D. McDermott(1985)
사람처럼 지능이 요구되는 기능들을 수행할 수 있는 기계를 만드는 기술	R. Kurzweil(1990)
지능형 행위를 계산 프로세스 측면에서 설명하고 에뮬레이션하는 연구 분야	J. S. Robert(1990)
사람이 더 잘하는 무엇인가를 어느 순간 컴퓨터가 할 수 있도록 하는 연구	E. Rich, K. Knight(1991)
지각, 추론 및 행동을 가능하게 하는 계산에 관한 연구	P. E. Winston(1992)
지능적 행위의 자동화와 관련된 컴퓨터 과학 분야	G. F. Luger, W. A. Stubblefield(1993)
인공적으로 만들어진 지능을 가지는 실제, 또는 그것을 만들고자 함으로써 지능 자체를 연구하는 분야	마쓰오 유타(2015)
인간의 인지능력(언어/음성/시각/감성 등)과 학습, 추론 등 지능을 구현하는 기술로 인공지능 SW/HW, 기초기술(뇌과학/산업수학 등)을 포괄	관계부처합동

출처 과학기술정책연구원, "인공지능 기술 전망과 혁신정책 방향-국가 인공지능 R&D 정책 개선방안을 중심으로," 정책연구 2018-13, 2018.

제해결 메커니즘을 정리한다. IV장에서 인공지능 세부 기술들이 자율주행 문제영역에 구현되고 적용되는 동향과 향후 발전 방향을 정리한 후 V장에서 결론을 맺는다.

## II. 자율주행

자율주행차는 운전자 또는 승객의 조작 없이 자동차 스스로 운행이 가능한 자동차로서, 자동차 스스로 사람의 인지, 판단, 제어 기능을 대체하여 운전하는 자동차를 말한다.

자율주행차 시스템의 구조는 구현 방식에 따라

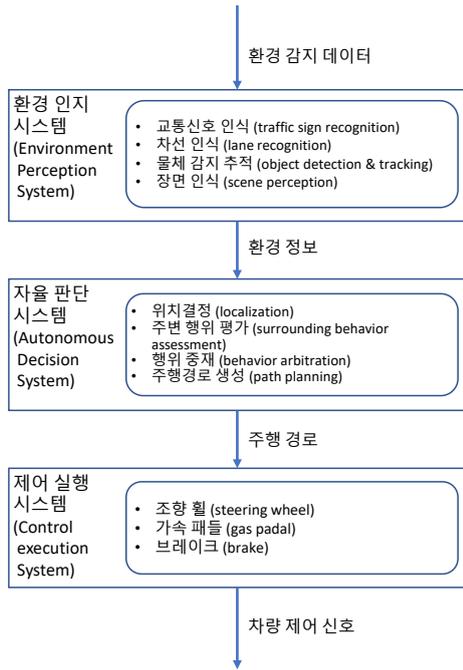


그림 1 자율주행차 개념 모듈 구조[4-7]

다양한데, 통상적으로 특정 임무(기능)를 담당하는 다수의 모듈이 자율주행 과정(단계)을 따라 파이프라인 방식으로 결합된 구조가 일반적이다. 이러한 구조는 거대하고 복잡한 문제를 더 작고 단순한 문제들로 분해하여 개별적으로 공략하는 전통적인 문제해결 접근법에 해당한다[4-7].

그림 1이 파이프라인 방식 시스템의 개념적인 모듈 구조를 보여준다. 인지 단계에서는 레이더 (radar), 라이더(lidar), 카메라(camera), 초음파 센서 (ultrasonic sensors) 등을 통해 주변 환경의 데이터를 수집하고 이를 이용하여 차량, 보행자, 도로, 차선, 장애물 등을 인지한다. 판단 단계에서는 앞 단계에서 인지된 결과를 바탕으로 주행 상황을 인식하고, 주행경로를 탐색하며, 차량/보행자 충돌 방지, 장애물 회피 등을 판단하여 최적의 주행조건(경로, 속도 등)을 결정한다. 제어 단계에서는 인지, 판단 결과를 바탕으로 차량 주행 및 움직임과

관련된 구동계, 조향계를 제어한다.

### 1. 환경 인지

환경 인지는 환경 감지 데이터를 처리하여 자율주행을 위해 필요한 정보를 추출하는 과정에 해당한다. 환경 데이터는 레이더, 라이더, 카메라, 초음파 센서 등 다양한 센서를 통해 수집할 수 있는데, 최근의 기술 동향은 저가의 카메라를 주요 센서장치로 사용한다. 카메라가 다른 센서장치에 비해 저렴하다는 장점뿐만 아니라 감지 속도가 빠르고 획득되는 정보량이 많기 때문이다[4-7].

교통표지 및 교통신호 인지는 자율주행차에게 매우 중요하다. 교통표지를 인식할 수 있다면 자율주행차 운행과 안전에 도움이 되는 정보를 얻을 수 있기 때문이다. 마찬가지로 교통신호를 인지하는 것은 교차로나 횡단보도와 같은 곳에서 교통규칙을 준수하게 함으로써 치명적인 사고를 방지하기 때문에 매우 중요하다.

자율주행차가 차선을 정확하게 인식하는 것은 주행차로 이탈 경고, 충돌방지, 주행경로 생성 및 다른 많은 하위 작업에 필수적이기 때문에 중요한 기능이다. 차선의 경우 자율주행차가 주행 중인 차로뿐만 아니라 다른 차로의 주행 방향, 병합 등 도로의 구조적 의미를 이해하는 데도 필수적이라는 측면에서도 매우 중요하다.

물체 감지는 관심 물체의 위치와 크기를 알아내는 것을 말하며, 감지의 대상은 신호등, 교통표지판 등 정지된 물체뿐만 아니라 다른 자동차, 보행자, 자전거 등 움직이는 물체를 포함한다. 더 나아가 자율주행차의 주행을 가로막는 장애물을 식별하는 것은 근본적인 기능에 해당한다.

빠른 속도에서의 주행, 복잡한 시내 변화가에서의 주행과 같이 높은 수준의 자율주행을 위해서는

물체, 장애물의 위치를 알아내는 것만으로는 부족하다. 움직이는 물체의 방향과 속도를 기반으로 해당 물체의 움직임을 예측할 수 있어야 한다. 즉 시간에 따른 궤적을 인지하는 것이 주행경로 생성, 충돌 회피 등에 필요하다.

주행전략 결정, 주행경로 생성 등에 있어 높은 수준의 자율주행을 위해서는 자율주행차가 놓인 장소의 의미적 정보(Semantic Information)를 인지할 필요가 있다. 예를 들어 학교 주변에서는 속력을 줄이고, 눈이 오는 도로에서는 미끄럼 방지 기능을 켜는 등의 결정을 할 수 있어야 한다. 더 나아가 장면에서 도로, 차선, 보도, 가로수, 건물 등에 해당하는 구역을 구분하여 분할할 수 있다면 더욱 정확한 장면 인지가 가능하다.

## 2. 자율 판단

자율 판단은 환경 인지를 통해 추출한 환경 정보를 바탕으로 자세, 충돌 예측, 주행경로 등 자율주행차 제어 신호 생성을 위해 필요한 여러 가지 사항을 결정하는 단계이다[4-7].

우선, 자율주행차는 자신의 자세(Pose)를 정확히 알고 있어야 한다. 즉 현재의 위치와 방향을 알아야, 예를 들어 자신의 현재 주행차로를 판단하고 올바른 주행경로를 생성할 수 있다.

안전한 주행을 위해서는 주변의 다른 운전자나 보행자의 의도를 판단하고 예측해야 한다. 이러한 예측을 이용하여 안전한 주행경로나 속도를 결정할 수 있기 때문이다. 정확한 예측을 위해서는 사람의 일반적인 특성, 다른 운전자의 주행 스타일 등이 고려되어야 한다.

자율주행차는 자신의 자세와 주변의 특성에 관한 판단과 예측을 통해 세부 자율주행 행위들을 조정하게 된다. 즉 장애물의 위치, 자신의 자세 및

속도, 주변 다른 자동차의 움직임 예측, 보행자의 궤적 등을 종합적으로 고려하여 차선 유지, 추월, 양보, 차선 변경, 교차로 회전, 긴급 제동 등에 관련된 시점, 우선순위 등 다양한 사항을 결정해야 한다.

위와 같은 판단은 최종적으로 주행경로 생성으로 모아진다. 주행경로 생성의 목적은 최종 목적지까지 안전하고 실패 없이 도달하는 방법을 매 순간 제공하는 것이다. 즉, 주행경로와 함께 속도 등 움직임(Motion) 관련 정보를 생성한다.

## III. 인공지능

### 1. 세부 인공지능 기술

인공지능 기술에는 다양한 세부 기술들이 포함된다. 한편, 인공지능은 일반적으로 인간의 사고와 지능, 행동 양식의 핵심능력을 모방한 모든 기술을 의미한다. 그런데 이러한 인공지능의 정의에 포함되는 사고, 지능과 같은 개념의 의미가 모호하고 추상적이기 때문에 인공지능의 세부 기술들을 구분하고 분류하는 방식 또한 다양하다[3].

인공지능 기술의 개발 및 발전 동향을 파악하기 위해서는 인공지능 기술들이 다양한 실제 문제에 어떻게 응용되는지를 살펴보아야 하는데, 이를 위해서는 응용과는 독립적인 기술 분류체계를 마련해야 한다. 본 고에서는 과학기술정책연구원에서 발표한 분류체계[3]를 채택하여 이 분류체계의 세부 기술들이 자율주행 문제를 해결하기 위해 응용되는 내용을 살펴봄으로써 기술 동향을 정리한다.

과학기술정책연구원 인공지능 기술 분류체계는 그림 2와 같이 3개의 위계로 구성된다. 우선 Level 1(대분류)에서는 지능형 에이전트가 환경과 상호작용하는 과정을 모사해, 외부 환경의 정보를 인식하는 단계와 이를 분석하는 단계로 구분하였다.

Level 1	2. 분석									
	Level 2	분류 (classification)		군집화 (clustering)		생성 및 의사결정 (generation/decision-making)				
		기계학습 기계학습	기계학습 신경망 (딥러닝)	기계학습 기계학습	기계학습 신경망 (딥러닝)	기계학습 신경망 (딥러닝)	기계학습 신경망 (딥러닝)	신경망 (딥러닝)		
2. 인식	이미지·영상	SIFT, SURF	Linear Regression, SVM, HMM, GMM, NNC	CNN, CNN+(a), GAN	Rule-based model	K-means, Fuzzy-C means, GMM, CRF	CNN, GAN	Wiener filter	GMM, CRF	CNN, RNN, DQN, DRQN, A3C, TRPO, DDPG, GAN
	신호	DTW, VQ, Quantization	Linear Regression, GPR, SVM, HMM, Random Forest	RNN, RNN+(a), DBN	N-gram model, Rule-based model	K-means, GMM, HMM, Kalman filter	RNN (LSTM), VAE, GAN	Time-series model	ADP, TD-learning, Kalman filter, Particle filter	DQN, DRQN, A3C, TRPO, DDPG, CNN, RNN
	텍스트·언어	Dictionary-based model	HMM, LDA, MNB	CNN, RNN, RNN+(a), RNNLM	Rule-based model	LSA, PLSA	CNN, RNN, GAN	LP, GA	Q-learning	RNN, GAN, VAE

주: Abbreviation: SIFT(Scale Invariant Feature Transform), SURF(Speeded-up Robust Features), DTW(Dynamic Time Warping), VQ(Vector Quantization), SVM(Support Vector Machine), HMM(Hidden Markov Model), GMM(Gaussian Mixture Model), LDA(Latent Dirichlet Allocation), MNB(Multinomial Naive Bayes), CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), LSTM(Long Short Term Memory Model), GAN(Generative Adversarial Network), DBN(Deep Belief Network), RNNLM(Recurrent Neural Network Language Modeling), LSA(Latent Semantic Analysis), PLSA(Probabilistic Latent Semantic Analysis), VAE(Variational Auto-Encoder), LP(Linear Programming), ADP(Approximate Dynamic Programming), GA(Generic Algorithm), CRF(Conditional Random Field), DRQN(Deep Recurrent Q Networks), A3C(Asynchronous Actor-Critic Agents), TRPO(Trust Region Policy Optimization), DDPG(Deep Reinforcement Learning), DQN(Deep Q Network)

출처 과학기술정책연구원, “인공지능 기술 전망과 혁신정책 방향-국가 인공지능 R&D 정책 개선방안을 중심으로,” 정책연구 2018-13, 2018.

그림 2 과학기술정책연구원의 알고리즘 관점에서의 인공지능 기술 분류체계[3]

Level 2(중분류)는 대분류별 목적에 따라 각각 세 개의 분류 기준을 구성하였다. 인식 단계에서는 인식하고자 하는 데이터의 속성에 따라 이미지·영상, 신호, 텍스트·언어로 구분하였으며, 분석 단계는 통상적 목적과 방식에 따라 분류(Classification), 군집화(Clustering), 생성 및 의사결정(Generation/Decision Making)으로 구분하였다. 그리고 마지막으로 Level 3(소분류)에서는 각 분류 기준별로 해당하는 대표 알고리즘들을 표기하였는데, 비기계학습과 기계학습을 구분하였고, 기계학습은 최근의 연구 동향을 반영해 신경망(딥러닝)을 이용한 기술과 그렇지 않은 기술로 유형화하였다.

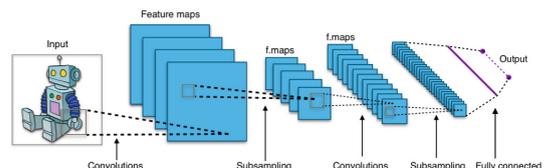
## 2. 주요 심층학습 기술

이 절에서는 앞 절에서 분류한 세부 인공지능 기술 중 심층학습을 기반한 인공지능 기술들의 특성을 살펴본다.

### 가. 합성곱 신경망

합성곱 신경망(CNN: Convolutional Neural Network)은 주로 이미지 인식에 응용되는 순방향 심층 신경망이다. CNN은 인접한 층과 층 사이에 특정한 유닛만이 결합을 갖는 특별한 층을 가지며, 이러한 층에서는 합성곱(Convolution)과 풀링(Pooling)이라는 이미지 처리의 기본적인 연산이 일어난다. 이는 사람의 시각 피질에 대한 신경과학적 지식으로부터 힌트를 얻어 만들어졌는데, 개별 피질 뉴런 중 단순세포(Simple Cell)는 감수영역(Receptive Field)으로 알려진 시야의 제한된 영역에서만 자극에 반응하고 출력층의 복잡세포(Complex Cell)는 중간층의 단순세포가 하나라도 활성화되면 활성화되는 것에 해당한다. 그림 3은 이러한 시각 피질을 모형화한 2층 구조의 신경망을 보여준다.

CNN은 일반 신경망보다 2차원 이상의 데이터를 입력받아 학습시키기 쉽고 고차원 데이터를 비교적 적은 매개변수를 이용하여 훈련시킬 수 있는 특징을 가진다. CNN을 이용한 영상 분류는 다른 영상 분류 알고리즘보다 상대적으로 전처리를 거



출처 Wikipedia, [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)

그림 3 합성곱 신경망의 구조

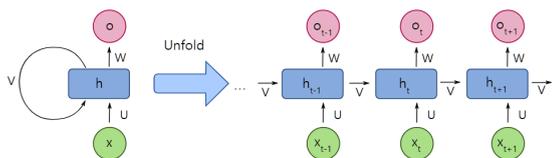
의 사용하지 않는다. 이는 신경망이 기존 알고리즘에서 수작업으로 제작된 필터를 스스로 학습한다는 것을 의미한다. 또한, 공유 가중치 구조와 변환 불변성 특성에 기초하여 변이 불변 또는 공간 불변 인공 신경망으로도 알려져 있다.

**나. 재귀 신경망**

재귀 신경망(RNN: Recurrent Neural Network)은 일반적인 인공 신경망과 달리 연속열 데이터를 분류하거나 예측하는 데 응용된다. 연속열 데이터란 각각의 요소에 순서가 있는 모임으로 주어지는 데이터를 말하는데 음성, 동영상, 텍스트 등이 데이터 요소에 순서가 있는 연속열 데이터의 예이다. 연속열의 길이는 일반적으로 가변적이다.

연속열 데이터를 다루는 추정 문제의 예로, 문장이 일부 주어졌을 때 다음에 출현할 단어를 예측하는 문제를 들 수 있다. 즉, 한 문장이 t번째의 단어까지 주어졌을 때 t+1번째 단어를 예측하는 것이 목표다. M개의 단어로 구성된 사전으로부터 각 단어를 레이블  $l = 1, 2, \dots, M$ 으로 나타내기로 할 때, 주어진 문장의 t번째 단어를 변수  $x_t$ , 예측할 단어를  $o_t$ 라 하면, 입력 연속열  $x_1, x_2, \dots, x_t$ 로부터 출력 연속열  $o_1, o_2, \dots, o_t$ 를 예측하는 문제로 볼 수 있다.

위 예에서, 문장은 단어의 자유로운 조합이기 때문에 무수히 많은 조합이 가능하지만, 실제로



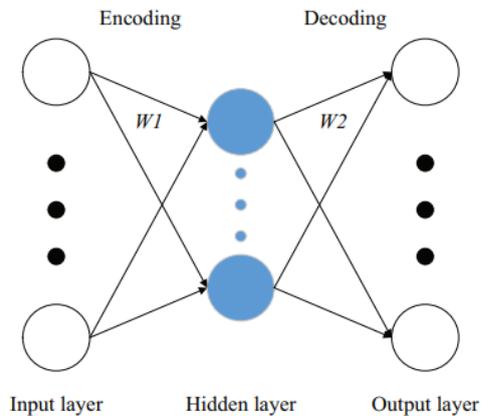
출처 Wikipedia, [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network)

그림 4 재귀 신경망의 구조

는 각 단어가 이전 단어의 연속열에 강하게 영향을 받는다. 즉 인간이 구사하는, 올바른 뜻을 가지는 문장은 단어 간의 의존 관계, 즉 문맥을 가진다. RNN은 그림 4와 같이 내부에 방향이 있는 순환경로를 가지며, 이러한 구조 덕분에 정보를 일시적으로 기억하고 그에 따라 반응을 동적으로 변화시킬 수 있다. 즉, 연속열 데이터 안에 존재하는 ‘문맥’을 포착하고 이를 바탕으로 이전에 입력된 데이터와 지금 입력된 데이터를 동시에 고려해 출력값을 결정한다.

**다. 자기 부호화기**

카메라나 라이다를 통해 영상을 획득하고 획득된 영상을 인식하여 환경의 상태를 인지하는 응용에서, 정확한 상황인지를 위해서는 고해상도의 이미지나 대용량의 포인트 데이터를 처리해야 한다. 그러나 입력 데이터의 크기(차원)가 커질수록 학습과 예측에 필요한 계산량이 폭발하는 단점이 있다. 자기 부호화기(AE: Auto-Encoder)는 이러한 문제에 적용할 수 있는 기술로, 데이터에 대한 효율적



출처 J. Ni et al., "A Survey on Theories and Applications for Self-Driving Cars Based on Deep Learning Methods," Applied Sciences, vol. 10 no. 8, 2020.

그림 5 자기 부호화기의 구조[5]

인 압축을 신경망을 통해 자동으로 학습하는 모델이다. 즉, 데이터의 차원을 줄이고 특징(Feature)을 자동으로 학습한다.

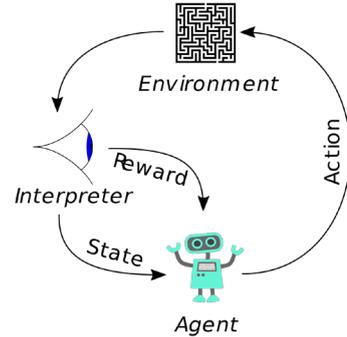
AE는 그림 5와 같이 크게 부호화기(Encoder)와 복호화기(Decoder)로 구성된다. 부호화기는 높은 차원의 원본 데이터를 작은 차원의 벡터 표현으로 변환함으로써 원본 데이터의 특징을 압축하는 역할을 한다. 복호화기는 반대로 작은 차원의 벡터 표현으로부터 원본 데이터를 복원하는데, 이를 통해 벡터 표현의 압축 품질을 평가하고 이를 바탕으로 학습을 진행할 수 있도록 한다.

우리의 관심사는 이렇게 훈련된 신경망에 어떤 데이터를 입력했을 때 중간층의 출력, 즉 입력 데이터의 부호(Code)다. 이를 이용하여 차원 폭발 문제에 대응할 수 있고, 잡음이나 불필요한 특징들을 걸러내고 중요 특징만을 학습할 수 있기 때문이다.

#### 라. 심층 강화학습

심층 강화학습(DRL: Deep Reinforcement Learning)은 심층학습과 강화학습을 결합한 기술이며, 최근 다양한 제어 문제에 적용되고 있다. 강화학습이란 에이전트(예, 로봇)가 주변 환경과 상호작용하면서 자신의 행동을 학습하는 기술이다. 이러한 기술에 대용량 데이터를 활용해 기존 기법들보다 압도적인 성능으로 학습하는 능력을 갖춘 심층학습 기법을 결합하여 기존에는 해결할 수 없었던 복잡한 문제에 대해서도 해답을 찾을 수 있게 되면서 주목받기 시작했다.

강화학습은 마르코프 결정 과정(Markov Decision Process)이라는 모델을 통해 형식화되는데, 예를 들어 로봇이 다양한 센서를 통해 현재 환경 상태를 인지하고 하나의 행동을 함으로써 환경으로부터 보상을 받고, 환경은 다음 상태로 전이된다. 강화학습은 이러한 시행착오 과정을 통해 하나의 최적



출처 Wikipedia, [https://en.wikipedia.org/wiki/Reinforcement\\_learning](https://en.wikipedia.org/wiki/Reinforcement_learning)

그림 6 강화학습 구조

정책을 학습하게 되는데, 정책은 각 상태에서 수행할 행동을 돌려주는 함수라 생각할 수 있으며, 최적 정책은 이러한 정책 중 누적 보상이 최대가 되도록 하는 정책이다. 그림 6이 개념적 강화학습 구조를 보여준다.

## IV. 자율주행 인공지능 기술 동향

이 장에서는 III장에서 정리한 심층학습 기반의 최신 인공지능 기술(모델)을 중심으로 II장에서 정리한 자율주행 응용 도메인에 적용되는 인공지능 기술의 동향을 정리한다. 몇 가지 언급해야 하는 사항은 첫째, III장에서 정리한 기술들은 기본모델에 해당하는 것으로 실제로는 기술별로 다시 다양한 변형과 발전된 기법들이 존재한다는 점이다. 둘째, 자율주행 도메인에 대한 인공지능 기술 동향을 정리하지만 이러한 내용은 비슷한 문제를 해결해야 하는 다른 도메인으로 자연스럽게 확장될 수 있다는 점이다.

### 1. 환경 인지 인공지능 기술

교통표지 인식에 심층학습 기반 기술이 최근

현저한 성능향상을 보여주고 있다. 기본적으로 CNN 기반의 알고리즘을 적용되는데, 다중 네트워크를 사용하여 우선 입력된 이미지 중 교통표지가 포함된 이미지를 고르고, 또 다른 네트워크를 이용하여 교통표지 영역을 추출한 후 마지막으로 교통표지를 분류(Classification)하는 방식을 사용한 다[10]. 즉, 이러한 방식에서는 CNN의 분류 정확도가 주요 역할을 한다. 교통신호 인식에 대해서도 기존 방식의 이미지 전처리나 의미적 장면 분할(Semantic Scene Segmentation)을 통해 신호등을 인식하고 CNN 네트워크를 통해 신호를 분류한다[11].

차선 인식은 전통적으로 원본 이미지에 대해 색 조절, 조명 정규화 등을 거친 후 컬러, 명암, 그라디언트 등의 정보를 다양한 필터를 이용하여 추출하여 차선을 구분하는 식으로 수행되었다. 최근에 CNN 네트워크를 이용하여 입력 이미지에서 차선에 해당하는 픽셀을 직접 분류하는 연구가 진행되었고[12], 또 다른 접근방식에서는 CNN 네트워크를 이용해 다양한 수준에서 특징(Feature)을 추출하고 추출된 특징에 분류 트리(Decision Tree), SVM(Support Vector Machine) 등 전통적인 기계학습 기법을 적용하거나[13], 심층학습 기술인 RNN 네트워크를 적용하여 차선을 구분하는 연구가 보고되었다[14].

물체 인식 기술 연구는 거의 50년 전부터 시작되었지만 1990년대 말 2000년대 초반에 와서야 자율주행차에 적용할 수 있는 정도의 성능에 도달하였고, 2012년 AlexNet[15]에 의해 심층학습 기법이 대세를 이루게 되었다. 일반적으로 심층 CNN을 통해 입력 이미지로부터 현저하게 축소된 해상도의, 그래서 짧은 시간에 처리 가능한 수준의 특징이 추출되고 추출된 특징이 완전결합 순방향 네트워크를 통해 물체를 분류하거나 바운딩 박스(Bounding Box) 픽셀, 깊이를 추정하는 구조를 가진다[16].

RNN을 적용하여 물체의 깊이 정보를 직접 추정하는 연구도 있는데, 이 경우 연속적인 스테레오 비디오 데이터가 입력되고 각 프레임 단위의 깊이지도(Depth-map)가 바로 출력되는 방식이다[17].

주변 물체의 미래 움직임, 즉 궤적, 속도, 방향 등을 예측하는 물체추적(Object Tracking)에 대해서는 일반적으로 연속적인 물체 인식 결과에 베이스 필터(Bayes Filter), 칼만 필터(Kalman Filter)를 적용하여 부드러운 동역학 모델을 계산하는 방식이 적용된다. 최근에 CNN, RNN 기반 방법으로 입력 이미지에서 실시간으로 물체를 인식하고 추적하는 연구가 이루어지고 있다[18].

심층학습 기술은 특히 이미지로부터 장면을 분류하거나 장면을 인식하는 데 명확한 이점을 가진다. 이미지 기반 물체인식 기술의 발전에 따라 최근에 심층학습을 자율주행의 장면이해(Scene Understanding)에 적용하기 위한 이론적 실험적 연구가 많이 이루어지고 있다. 대부분의 연구에서는 다중 해상도 CNN을 이용하여 서로 다른 규모에서의 시각적 구조(특징)를 학습하고 이러한 구조를 이용하여 다양한 규모에서의 주행 장면 이해에 적용한다[19]. 한편, 구획(Segmentation), 분류(Classification) 그리고 문맥 통합(Context Integration)을 LSTM(Long Short-Term Memory) RNN을 이용하여 한꺼번에 수행하고 주변의 문맥 정보와 식별된 레이블(물체, 차선, 신호등 등) 간의 공간적 의존관계를 학습하는 연구가 보고되었다[20].

## 2. 자율 판단 인공지능 기술

자율주행차가 자신의 위치와 자세를 추정하는 기능은 기본적으로 GPS(Global Positioning System)를 이용하여 구현할 수 있지만 최근 심층학습 기술을 이용해 시각 데이터로부터 주변 환경 정보를

추출하여 자율주행차의 위치와 자세를 추정하는 기술개발이 이루어지고 있다. 비주얼 오도메트리(Visual Odometry)라 불리는 로컬라이제이션 기술은 연속적인 비디오 프레임 간에 특징적 랜드마크(Keypoint Landmark)를 매칭시킴으로써 이전 프레임 대비 현재 프레임에서의 자율주행차의 위치와 자세를 추정한다[21]. 심층신경망을 특징점 이상치(Keypoint Outlier)를 구분하도록 학습시킴으로써 특징적 랜드마크 감지 정확도를 향상시키는 방식으로 응용될 수 있다. 또한, 이미지 데이터로부터 직접적으로 카메라 자세(Pose), 즉 자율주행차의 자세를 추정하는 시스템도 개발되었다[22].

주변 주행차의 미래 움직임을 예측하는 문제는 전통적으로 은닉 마르코프 모델(HMM: Hidden Markov Model)이나 베이지안 네트워크(Bayesian Network)를 이용하여 해결하고 있는데, 최근 심층 학습을 이용해 장면 흐름(Scene Flow)을 학습함으로써 전통적인 접근방식을 대체하려는 연구가 진행되었다. 이 연구에서는 심층신경망이 점유 그리드(Occupancy Grid)상에서 서로 다른 시점 간에 일치되는 위치를 추적하도록 훈련된다[22].

주변 운전자의 운전 스타일을 판단하는 데도 심층 학습이 응용된다. 우선, 심층 자기 부호화기를 이용하여 다양한 운전자의 특징적인 운전자 특성을 비지도 학습 방식으로 추출하고 이를 이용하여 판단 대상 운전자의 유형을 군집화 알고리즘으로 분류할 수 있다[23,24].

자율주행차가 장애물의 위치, 자신의 자세 및 속도, 주변 다른 자동차의 움직임 예측, 보행자의 궤적 등 주변 환경과 상황을 종합적으로 고려하여 차선유지, 추월, 양보, 차선 변경, 교차로 회전, 긴급 제동 등에 관련된 시점, 우선순위 등 다양한 사항을 결정하는 것은 매우 복잡하고 어려운 작업이다. 기존 연구에서는 이를 위해 임무 계획, 행위 계획,

모션 계획 등으로 세분화하고 다양한 기술을 적용하였다[6].

최근 심층 강화 학습을 적용하여 다양한 상황에 대응 가능한 행동 정책을 학습하는 연구가 진행되고 있다[6,25,26]. 강화 학습은 사람이 직접 학습 데이터를 미리 작성할 필요 없이 학습 시스템이 환경과 상호작용을 하면서 시행착오를 통해 스스로 학습하기 때문에 자율주행차 도메인에 매우 적합한 기술이다. 또한 심층 학습의 도움으로 한 번도 경험하지 못한, 즉 훈련 데이터에 포함되지 않은 상황에 대해서도 학습된 결과를 일반화할 수 있다. 그러나 많은 시행착오를 위해 반드시 시뮬레이션 환경이 필요하며, 시뮬레이션 환경의 실환경 모사 정확도에 큰 영향을 받는 단점이 있다.

## V. 결론

지금까지 자율주행차 도메인에 대해 심층 학습을 중심으로 인공지능 기술의 동향을 살펴보았는데, 특히 CNN, RNN, AE, DRL 네 가지 심층 학습 기술(모델)이 자율주행차 도메인의 다양한 문제에 적용되는 기술 동향을 그림 7과 같이 요약할

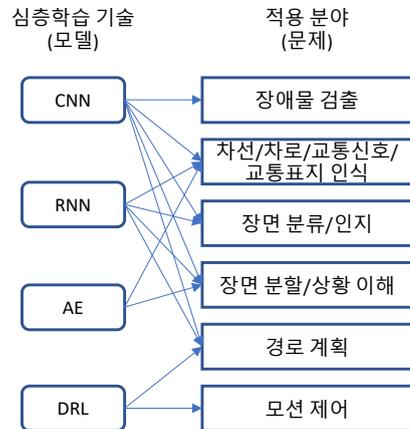


그림 7 심층 학습 기술 적용 구조

수 있다.

심층학습 기술이 이미지 인식에 있어 놀라운 성능향상에 힘입어 최근 자율주행차 도메인에 새로운 가능성과 큰 발전을 제공하고 있다. 그러나 여전히 개선되고 해결해야 하는 많은 도전적인 문제들이 남아 있다. 첫째, 학습 데이터로 사용할 예제(Example)에 관련된 문제인데, 학습 결과의 정확도와 일반화 성능을 보장할 수 있는 우수한 품질의 예제를 대량으로 마련하는 것이 심층학습을 자율주행차에 적용하는 데 중요한 문제 중 하나이다. 둘째, 현재 심층학습 기술은 주로 이미지를 기반하는데 실제 주행 중 획득되는 이미지는 기후, 조명, 오염 등 다양한 원인으로 품질이 나빠지거나 학습할 때 사용했던 이미지와 차이를 보일 수 있다. 이럴 때도 안정적인 추론(추정) 성능을 보장하는 것이 문제이다. 또 다른 문제로 처리 속도 문제가 있다. 실세계는 일반적으로 매우 큰 차원을 가지는 연속 상태 공간 학습 데이터로 대응되는데, 이러한 문제에 대해 추론을 수행하는데 아직 심층학습 기술은 사람의 인지, 판단, 제어 속도를 따라가지 못한다. 따라서 신속한 인지, 판단, 제어가 요구되는 위급상황에 적용할 수 있도록 개선되어야 한다.

위와 같은 문제에도 심층학습 및 관련 인공지능 기술이 향후 자율주행차를 실현하는 핵심 기술로 자리 잡을 것으로 예상된다. 또한, 현재 자율주행에 필요한 인지, 판단, 제어 등 단계를 기능별로 구분하여 소프트웨어를 개발, 운용하는 방식과는 달리, 심층학습 기반의 인공지능 기술에 의존해 전체 과정을 한번에 구현하는 엔드투엔드(End-to-End; 종단 간) 방식의 자율주행 기술도 활발히 연구개발되고 있다.

#### 용어해설

**Lidar(라이더)** 라이더라는 명칭은 전파 대신에 빛을 쓰는 레이더를 뜻하는 것으로 라이더는 레이저 펄스를 발사하고, 그 빛이 주위의 대상 물체에서 반사되어 돌아오는 것을 받아 물체까지의 거리 등을 측정함으로써 주변의 모습을 정밀하게 그려내는 장치이다.

**AlexNet** AlexNet은 ImageNet 데이터 셋을 이용하여 이미지 분류 성능을 겨루는 대회인 ILSVRC에서 2012년에 압도적 성능으로 우승한 CNN 모델이다. 개발자인 Alex Krizhevsky의 이름을 따서 명명되었다.

**Kalman Filter(칼만 필터)** 잡음이 섞여 있는 기존의 관측값을 최소 제곱법을 통해 분석함으로써 일정 시간 후의 위치를 예측할 수 있도록 하는 최적의 수학적 계산 과정

#### 약어 정리

AE	Auto-Encoder
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DRL	Deep Reinforcement Learning
GPS	Global Positioning System
HMM	Hidden Markov Model
IoT	Internet of Things
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network

#### 참고문헌

- [1] 최진철, 손영성, "인공지능 기반 자율 사물 개발 동향과 발전 전망," 주간기술 동향, 1937호, 2020.
- [2] Gartner, "Top 10 Strategic Technology Trends for 2019," G00374252, 2018.
- [3] 양희태 외, "인공지능 기술 전망과 혁신정책 방향-국가 인공지능 R&D 정책 개선방안을 중심으로," 정책연구 2018-13, 2018.
- [4] KDB 산업은행 경제연구소, "자율주행차 국내외 개발 현황," 이슈분석 제771호, 2020.2.
- [5] J. Ni et al., "A Survey on Theories and Applications for Self-Driving Cars Based on Deep Learning Methods," Applied Sciences, vol. 10 no. 8, 2020.
- [6] S. Grigorescu et al., "A survey of deep learning techniques for autonomous driving," Journal of field robotics, vol. 37 no. 3,

- Apr. 2020.
- [7] E. Yurtsever et al., "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," *IEEE Access*, vol. 8, 2020.
- [8] 스투어드 러셀, 피터 노빅, "인공지능: 현대적 접근방식," 제이펍, 2016.
- [9] Tractica, "Artificial intelligence for enterprise applications," 2015.
- [10] H. Xu and G. Srivastava, "Automatic recognition algorithm of traffic signs based on convolution neural network," *Multimedia Tools and Applications*, vol. 79, Jan. 2020.
- [11] E. Lee and D. Kim, "Accurate traffic light detection using deep neural network with focal regression loss," *Image and Vision Computing*, vol. 87, July. 2019.
- [12] J. Kim, J. Kim, G. Jang and M. Lee, "Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection," *Neural Networks*, vol. 87, Mar. 2017.
- [13] V. John et al., "Real-time road surface and semantic lane estimation using deep features," *Signal, Image and Video Processing*, vol. 12, Mar. 2018.
- [14] D. Neven et al., "Towards End-to-End Lane Detection: an Instance Segmentation Approach," arXiv: 1802.05591, 2018.
- [15] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.* 2012.
- [16] M. Mancini et al., "J-MOD2: Joint monocular obstacle detection and depth estimation," *IEEE Robot. Autom. Lett.* 2018.
- [17] Y. Zhong, H. Li and Y. Dai, "Open-world stereo video matching with deep RNN," in *Proc. Eur. Conf. Comput. Vis.* Sep. 2018.
- [18] B. Huval et al., "An empirical evaluation of deep learning on highway driving," Apr. 2015, arXiv:1504.01716
- [19] L. Wang et al., "Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs," *IEEE Trans. Image Process.* 2017.
- [20] W. Byeon et al., "Scene labeling with LSTM recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Jun. 2015.
- [21] D. Barnes et al., "Driven to Distraction: Self-Supervised Distractor Learning for Robust Monocular Visual Odometry in Urban Environments," in *Proc. IEEE Int. Conf. on Robotics and Automation*, 2018.
- [22] A. K. Ushani and R. M. Eustice, "Feature Learning for Scene Flow Estimation from LIDAR," in *Proc. Conf. Robot Learning*, vol. 87, Oct. 2018.
- [23] E. Yurtsever, C. Miyajima and K. Takeda, "A traffic flow simulation framework for learning driver heterogeneity from naturalistic driving data using autoencoders," *Int. J. Automat. Eng.* vol. 10 no. 1, 2019.
- [24] K. Sama et al., "Driving feature extraction and behavior classification using an autoencoder to reproduce the velocity styles of experts," in *Proc. Int. Conf. Intell. Transp. Syst.* Nov. 2018.
- [25] S. Shalev-Shwartz, S. Shammah and A. Shashua, "Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving," 2016.
- [26] A. I. Panov, K. S. Yakovlev and R. Suvorov, "Grid Path Planning with Deep Reinforcement Learning: Preliminary Results," *Procedia Computer Science*, vol. 123, 2018.