

# 자율주행차량의 인공지능 기술 동향

박재호

데일리인텔리전스 CTO

SF 소설과 영화에 단골손님으로 나오던 자율주행차량은 최근 기술 발전에 힘입어 더 이상 상상의 산물이 아니라 일상화된 현실로 자리잡고 있다. 자율주행차량은 주변 환경을 실시간으로 파악하고 차량 스스로 결정을 내려 운행이 가능하므로 기존에 볼 수 없던 새로운 혁신을 이끌어내어 차량 업계뿐만 아니라 관련 산업의 지형도를 바꾸고 있다. 본 고에서는 자율주행차량의 센서와 CPU/GPU 등 하드웨어 구성 요소를 설명한 다음에 본격적인 소프트웨어 기술 발전을 머신러닝과 딥러닝을 중심으로 설명한다. 전통적인 영상처리의 문제점을 정리한 다음에 이를 극복하기 위한 다양한 기술을 비교하며, 학습 과정에 필요한 데이터의 중요성과 시뮬레이터 개발 동향도 소개한다.

## I. 서론

지난 5월에 열린 개발자 행사인 구글 I/O에서 알파벳 소속 웨이모의 키노트가 언론의 주목을 받았다[1]. 구글이 영상 분석과 음성 인식 분야에서 많은 연구를 진행해 왔지만 이를 어떻게 자율주행차량과 연계시켜 왔는지는 외부에 잘 알려지지 않았다. 하지만 이번 키노트에서 몇 가지 중요한 사항이 공개되었다. 캘리포니아 공과대학의 보행자 데이터를 대상으로 하는 보행자 인식률의 경우 초기 연구 단계에서는 평균 오류율이 26.2%였지만 지속적인 연구에 힘입어 100배 이상 개선했으며, 공사 현장 우회, 긴급 차량 양보, 평행 주차 시 다른 차량 배려와 같이 인간에 가까워지기 위해 딥러닝 기술을 예측, 계획, 매핑, 시뮬레이션에 이르는 전반적인 과정에 활용하고 있다. 복잡한 알고리즘을 사용해 학습과 시뮬레이션을 하려면 하드웨어와 소프트웨어 기반 구조가 튼튼하게 뒷받침되어야 하는데, TPU(Tensor Processing Unit),

\* 본 내용은 박재호 CTO(02-6105-7039, jayhpark@daylifg.com.)에게 문의하시기 바랍니다.

\*\* 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

구글 클라우드 플랫폼, 텐서플로를 중심으로 구글이 현재까지 개발한 다양한 기술을 토대로 25,000대에 이르는 가상 차량이 시뮬레이터에서 운행되고 있다. 폭우나 폭설과 같은 악천후에 대응하기 위한 여러 가지 기술도 머신러닝을 사용해 해결하고 있다는 키노트 마무리에서 자율주행차량이 현실에 더욱 가까워졌다는 사실을 확인할 수 있었다.

비록 최근에 발생한 인명사고로 인해 잠정적으로 자율주행차량 테스트를 중단하기는 했지만, 공유차량 서비스를 제공하는 우버에서도 자율주행차량의 하드웨어와 소프트웨어에 대해 집중적인 투자를 진행하고 있다. 사람의 도움 없이 차량이 연중무휴(24×7×365) 차량 배차 서비스를 제공한다면 기존 자동차 제조사와 대중교통 업계의 지형이 완전히 바뀌게 되며 이는 자동차가 처음 등장한 이후 지금까지 누구도 예상하지 못했던 근본적인 교통수단의 혁명이 도래함을 의미한다.

본 고에서는 하드웨어 관점에서 자율주행차량의 구성 요소(센서, CPU/(GP)GPU)를 소개한 다음에 소프트웨어 관점에서 영상 분석에 뛰어난 머신러닝과 딥러닝을 소개한다. 전통적인 영상 처리 기술의 한계점을 극복하는 새로운 기술인 딥러닝 중에서도 특히 이미지에 대한 객체 판별/분류와 검출에 가장 널리 사용되는 CNN 계열의 학습 방법을 소개한다. 또한, 데이터의 중요성과 시뮬레이터 기술의 발전 현황을 소개한다. 마지막으로 결론과 시사점으로 마무리한다.

## II. 자율주행차량의 구성 요소 소개

### 1. 실수가 없는 사람을 닮기 위한 노력

자율주행차량은 사람의 개입없이 처음부터 끝까지 차량 스스로의 판단만으로 운행이 가능해야 한다는 궁극적인 목표로 향해 가고 있다. 국제자동차기술자협회(Society of Automotive Engineers: SAE)의 자율주행 6단계 [표 1]에 따르면 자동화가 전혀 없는 0단계에서 어떤 상황에서도 완전자율주행이 가능한 5단계까지 나뉘지는데, 각 단계별로 풀어야 하는 문제가 존재한다. 여러 회사들이 제한된 자율주행이 가능한 3단계와 특정 상황을 제외한 나머지 상황에서 완전자율주행이 가능한 4단계를 놓고 치열한 경쟁을 벌이고 있지만, 최근에 자율주행 시험 차량 보급이 늘어나면서 여러 가지 한계점이 드러나고 있으며 이를 극복하기 위한 다양한 수단도 함께 등장하고 있다.

만일 실수가 없는 가상적인 사람이 운전석에 앉아있다고 가정할 때 완벽한 운전이라는 행

[표 1] 국제자동차기술자협회(SAE)의 자율주행 6단계

| 구분     | 0단계 | 1단계    | 2단계        | 3단계     | 4단계    | 5단계    |
|--------|-----|--------|------------|---------|--------|--------|
| 명칭     | 수동  | 운전자 보조 | 일부 자동화     | 조건부 자동화 | 높은 자동화 | 완전 자동화 |
| 운전자 개입 | 항상  | 항상     | 항상         | 요청에 따라  | 특정 상황  | 없음     |
| 관련 기술  | 없음  | ADAS   | ADAS       | ADS     | ADS    | ADS    |
| 관련 회사  |     |        | 테슬라(2.5단계) | 우버      | 웨이모    |        |

\* ADAS(Advanced Driver Assistance System), ADS(Automated Driver system)

&lt;자료&gt; NHTSA, “Automated Vehicles for Safety”

동은 어떤 식으로 이루어질까? 가장 먼저 출발지에서 목적지로 어떻게 갈지 경로를 정해야 한다. 지도를 보고 이어지는 도로를 탐색하면서 경유지나 교차로를 파악하는 방법을 사용한다. 실제 운전에서 들어갈 경우 현재 위치 파악이 중요하다. 각종 이정표와 주변 지형지물을 활용해 올바른 도로에 위치해서 올바른 차선을 타고 있는지 알아야 지속적으로 차량 제어가 가능해진다. 실제 주행 중에는 신호등과 표지판은 물론이고 다른 차량이나 보행자나 장애물을 감지하고 각각의 의도를 파악해야 한다. 신호등과 표지판을 보고 교통 법규를 지키는 동시에, 다른 차량이 차선을 바꾸는지 혹은 갑자기 정체로 인해 속력을 줄이는지를 주의 깊게 봐야 하며, 보행자가 정말 길을 건널지 혹은 멈춰 서 있을지 아니면 지금 느리게 길을 건너고 있는지 아니면 뛰어가고 있는지를 파악해야 한다. 시설물 보수 등으로 인해 차선이 통제될 경우에도 이를 인지해서 우회해야 한다.

자율주행차량의 운전 과정에서 드러나는 몇 가지 중요한 요소는 출발지에서 목적지까지 가는 방법을 아는 경로 설정, 내가 현재 있는 위치를 아는 측위, 도로 차선/교차로/신호등/표지판 상황을 아우르는 정밀 지도, 현재 주변에 위치한 장애물/차량/사람의 움직임과 의도를 아는 주변 상황인지로 정리할 수 있다. 이 모든 상황을 종합적으로 파악해서 스티어링 휠의 각도, 가속 페달 각도, 브레이크 압력을 조정하면 안전한 운전이 가능해진다. 이 모든 일이 원활하게 이루어지기 위해서는 자율주행차량의 눈과 귀로 볼 수 있는 센서와 두뇌로 볼 수 있는 CPU/(GP: General Purpose)GPU가 필요하다.

## 2. 센서: 자율주행차량의 눈과 귀

사람이 운전 도중에 눈과 귀를 사용해 운전에서 필요한 정보를 획득하듯이 자율주행차량도 현재 위치와 외부 환경을 인지하기 위해 센서를 사용한다. 센서 하나로 모든 정보를 알 수

없기에 여러 종류의 센서를 여러 개 장착하고 있다.

#### 가. GPS

차량의 정확한 위치를 파악하기 위해 사용하는 센서이다. 지구 주위를 돌고 있는 위성들로부터 수신한 전파를 사용해 송신된 신호와 수신된 신호의 시차를 측정하는 방법으로 주변 지형지물의 도움 없이도 실시간으로 차량의 정확한 위치를 계산한다. 하지만 고층 건물이 즐비한 도심, 도로 아래, 터널이나 지하 구간에서는 수신이 되지 않는 문제점이 있다. 따라서 차량의 속도 정보를 사용해 현재 위치를 추정하기 위한 추측항법 시스템이 결합된 형태로 단점을 어느 정도 보완해야 한다.

#### 나. 카메라

차선을 인식하고 주변 사물 형태와 색상을 인식하기 위해 사용하는 센서이다. 가장 범용적으로 사용 가능하며 해상도가 높은 정보를 획득하므로 핵심으로 자리 잡고 있다. 하지만 역광, 상향등, 안개와 같은 외부 조건에 취약하며 내용을 해석하는 과정에서 오류가 발생할 가능성이 있으므로 100% 신뢰할 수 없다는 문제점이 있다. 일반 카메라는 깊이를 인식하지 못하므로 렌즈를 두 개 사용하거나 카메라를 두 대 설치하는 방법으로 3차원 입체를 인식하게 만드는 방법도 사용한다.

#### 다. 라이다(Lidar)

레이저를 쏘아 돌아오는 초점 이미지와 시간을 계산하여 특정 지점의 위치를 파악하고 이를 하나로 모아 포인트 클라우드를 만들어내기 위해 사용하는 센서이다. 정확한 3차원 모델을 만들 수 있으므로 자율주행차량에서 사용되는 고해상도 지도를 만들기 위해서도 사용되며 주행 중에 주변 지물을 3차원으로 재구성해준다. 하지만 해상도가 낮으며 높은 비용과 비나 눈과 같은 기상 조건에 영향을 받는다는 문제점이 있다. 차량 제조사들과 부품 제조사들은 크기를 줄이고 비용을 대폭 낮춘 신형 라이다 시스템을 개발하고 있기 때문에 카메라만큼이나 일반화될 가능성이 높아지고 있다.

#### 라. 레이더

전자파를 쏘아 돌아오는 반사파를 계산하여 위치와 형상을 파악하기 위해 사용하는 센서이다. 주변 환경이나 기상 상황에 영향을 적게 받으며 움직이는 물체에 대해 상대적인 거리를 측정하기에 적합하다. 하지만 정확한 위치와 크기를 알아내기 어렵다는 문제점이 있다. 최근

에는 디지털화된 레이더와 소프트웨어 기술 발전에 힘입어 주파수 도약 기법을 사용하는 FMCW(Frequency Modulation Continuous Wave) 기술을 적용하여 사물의 정확한 위치와 형태를 영상으로 표현할 수 있는 합성 개구 레이더(Synthetic Aperture Radar: SAR)를 차량에 탑재하려는 움직임도 있다.

#### 마. 초음파

음파(50KHz 주변 초음파)를 쏘아 돌아오는 반사파를 계산하여 위치를 파악하기 위해 사용하는 센서이다. 근접거리에서 사물의 위치를 측정하기에 적합하다는 특성이 있다.

#### 바. 적외선

적외선 복사를 감지해서 사물의 위치와 형상을 파악하기 위해 사용하는 센서이다. 주변보다 높은 열을 발산하는 물체(차량 엔진과 사람)를 감지하므로 주변 환경이나 기상 조건에 영향을 적게 받는 동시에 정확한 위치와 열원의 크기를 파악할 수 있는 장점이 있다. 기술의 발전에 따라 전방의 열화상을 영상으로 변환하는 수준에 이르렀다(Forward Looking InfraRed: FLIR).

[표 2] 센서별 장애물 탐지 각도와 거리 비교

| 유형     | 탐지 각도                   | 탐지 거리         |
|--------|-------------------------|---------------|
| 라이다    | 360도(수평)/30도(수직)        | 120미터         |
| 광각 카메라 | 120도(수평)/90도(수직)        | 60미터          |
| 협각 카메라 | 30도(수평)/10도(수직)         | 250미터         |
| 레이더    | 60도(근거리 수평)/30도(원거리 수평) | 200미터         |
| 초음파    | 60도                     | 8 미터          |
| 적외선    | 광각/협각 카메라와 유사           | 광각/협각 카메라와 유사 |

<자료> 테슬라, “모든 차량을 위한 완전 자율 주행 하드웨어”

### 3. 두뇌: CPU와 GPU

인체의 여러 감각 기관에서 들어온 전기적인 신호를 처리하는 두뇌와 유사하게 여러 센서로부터 수집한 정보를 통합적으로 분석해서 처리하기 위해 자율주행차량은 CPU와 (GP)GPU를 사용한다. CPU는 외부 장치에서 들어온 정보를 해석하고 연산하고 상태를 기억하고 다시 외부 장치로 명령을 내리는 핵심 구성요소이다. 순차적인 직렬처리에 강한 CPU와는 달리 (GP)GPU는 복잡한 수치 연산을 빠르게 처리하기 위한 병행 작업 전용으로 만들어졌다.

## 가. CPU

차량에 특화된 CPU는 마이크로프로세서와 입출력 모듈이 하나로 결합된 형태의 MCU(Micro Controller Unit)가 가장 널리 사용된다. CPU 코어, 메모리, 각종 입/출력 장치, 플래시 메모리, EEPROM 등을 하나로 합친 형태이므로 부피도 작고 전력 소비도 적다는 장점이 있다. 하지만, 용도에 따라 미리 정해진 작업부하(workload)를 처리하므로 범용 CPU에 비해 클럭 주파수와 메모리 주소 공간이 제한적이다.

최근에는 요구 사항이 복잡해짐에 따라 중앙처리장치의 기본 단위인 코어 수를 여러 개로 늘린 멀티코어가 대세로 떠오르면서 단일 코어의 성능 한계를 극복하고 있으며, 신호처리 과정에서 하드웨어 가속이 필요한 부분에 대해서는 DSP까지 하나로 합친 제품까지 나오고 있다. 하지만 멀티코어로 인해 병렬 실행 기능이 개선되고는 있지만, 여전히 순차적 실행 기능이 우선이므로 복잡한 데이터를 계산하는 과정에서 부족한 부분이 존재한다.

## 나. (GP)GPU

기존 임베디드 부문에서는 빠른 신호처리를 목적으로 DSP(Digital Signal Processor)를 주로 사용했지만, 자율주행차량 특성상 아날로그 신호를 실시간으로 처리하는 능력 이외에 다양한 디지털 신호를 받아서 복잡한 계산을 수행하는 능력이 필요하다. 과거 GPU는 2차원 가속 기능을 제공하여 화면 출력을 빠르게 할 목적이었지만, 시간이 지남에 따라 3차원 그래픽을 위한 복잡한 벡터 연산 기능을 빠르게 할 목적으로 발전해 왔다. SIMD(Single Instruction Multiple Data) 방식의 병렬처리를 위해 많게는 수 천 개의 소형 코어를 장착한 GPU는 그래픽 작업뿐만 아니라 일반적인 수치연산 작업에도 적합하므로 범용(General Purpose) GPU로도 불린다.

예를 들어, 머신러닝 과정에서 많이 사용되는 벡터 연산을 생각해 보면,  $M \times N$  행렬과  $N \times P$  행렬을 곱할 때, 기존 CPU에서는  $M \times N$  행렬의 각 행의 원과  $N \times P$  행렬의 각 열의 원을 차례로 곱한 다음에 더해서  $M \times P$  행렬을 구하는 방식으로 프로그램을 작성해야 한다. 하지만 (GP)GPU를 사용할 경우 하드웨어 스레드를 사용하여 여러 곱을 동시에 수행한 다음 더할 수 있으므로 높은 성능을 발휘할 수 있다.

최근에는 모바일 장비에서 자연어 처리, 손글씨 인식, 얼굴 인식과 같은 기능이 요구됨에 따라 성능과 배터리 수명을 극대화하기 위해 CPU 내부에 GPU 기능이 탑재된 형태의 제품도 출시되고 있다. 예를 들어, 아이폰 X에 탑재된 A11 바이오닉 CPU의 경우 뉴럴 엔진이라 부르는 전용 뉴럴 네트워크 하드웨어를 포함하고 있으며, 페이스 ID와 같은 인공지능 기능을 구현하는 핵심 원동력이 되었다[2]. 이와 유사한 방식으로 인텔/모빌아이나 AMD에서는 자율주행

차량을 목표로 GPU 기능이 탑재된 저전력 고효율 CPU를 준비하고 있다.

#### 다. 학습 모델 생성을 위한 고성능 GPU와 전용 프로세서

자율주행차량을 위한 학습 모델을 만드는 작업은 방대한 데이터를 복잡한 신경망에 넣어서 처리하는 작업이 필요하므로 고성능 전용 프로세스가 있어야 시간을 단축할 수 있다. 크게 GPU 기술을 확장해서 사용하는 방법과 전용 프로세서를 직접 만들어 사용하는 방법이 있다. NVIDIA와 AMD에서 머신러닝을 위한 고성능 (GP)GPU를 개발했는데, NVIDIA가 만든 최신 GPU 코어인 TELS V100의 경우 심층신경망 학습 성능이 120테라플롭스를 넘어선 상황이며, NVLINK를 사용하여 V100 사이에 양방향 연결을 할 경우 300GB/s 대역폭을 제공하므로 서버 한 대에 여러 GPU를 장착하여 성능을 높일 수 있다[2]. 구글은 자사 클라우드 환경에 최적화하기 위해 ASIC(Application-Specific Integrated Circuit) 기술로 TPU를 개발하고 있으며, 최근 구글 I/O에서 발표된 3세대 TPU의 경우 45테라플롭스가 나오던 2세대 TPU에 비해 8배 정도의 성능을 발휘한다고 알려져 있다[3]. 구글, 아마존, 마이크로소프트에서는 엄청나게 많은 데이터를 빠르게 학습할 수 있게 GPU와 TPU를 탑재한 서버를 클라우드 형태로 제공하고 있다.

#### 4. 자율주행차량 하드웨어 사양 비교

지금까지 자율주행차량의 개별 하드웨어 구성요소를 소개했는데, 웨이모, 테슬라, 우버를 중심으로 전반적인 자율주행차량 하드웨어 사양을 비교하는 방법으로 통합된 모습을 살펴보

[표 3] 자율주행차량 하드웨어 사양 비교

|             | 웨이모                      | 테슬라               | 우버           |
|-------------|--------------------------|-------------------|--------------|
| 카메라         | 8대                       | 8대                | 7대           |
| 라이다         | 장거리(중앙)와 단거리(앞 두 모퉁이) 장착 | 없음                | 중앙 장착        |
| 레이더         | 5개(중앙과 앞뒤 네 모퉁이)         | 전방 레이더            | 10개(360도 탐지) |
| 초음파         | ?                        | 12개               | ?            |
| CPU/(GP)GPU | NVidia로 추정               | 모빌아이에서 NVidia로 변경 | Nvidia       |
| 소프트웨어       | 독자 개발                    | 모빌아이에서 독자 개발로 변경  | 독자 개발        |

<자료> Waymo Team, "Introducing Waymo's suite of custom-built, self-driving hardware," 2017. 2.

테슬라, "모든 차량을 위한 완전 자율 주행 하드웨어"

Heather Somerville, Paul Lienert, Alexandria Sag, "Uber's use of fewer safety sensors prompts questions after Arizona crash," Reuters, 2018. 3. 28.

겠다. 탑재된 소프트웨어에 대해서는 아직 구체적으로 발표된 사항이 없지만 하드웨어 사양 등으로 짐작해보면 대부분 딥러닝으로 만든 모델을 중심으로 독자적인 차량운행 소프트웨어를 사용하고 있을 것으로 추정된다.

### Ⅲ. 기계학습과 딥러닝

#### 1. 전통적인 영상처리의 한계

자율주행차량에서 가장 중요하게 처리해야 하는 데이터는 영상과 밀접한 관련을 맺고 있다. 앞서 설명한 센서 중에서 라이다, 카메라, 적외선(특히 FLIR), 레이더(아직은 차량용으로 실용화되지 않았지만 특히 SAR)에서 수집되는 영상 정보를 분석해야 차량 주변에 위치한 여러 가지 사물(차량, 방해물)과 사람에 대한 위치와 이동 상태를 파악할 수 있다.

전통적인 영상처리 방식에 따르면 화면에서 원하는 사물을 찾기 위한 알고리즘을 개발한 다음 실시간으로 프로그램을 동작시켜 결과를 얻는다. 어떤 사물을 특징할 수 있는 특징(feature)을 추출하여 색상, 크기, 형태를 토대로 분석하는 방식이 대표적이다. 하지만 오랫동안 연구와 실험을 계속했음에도 불구하고 다음과 같은 근본적인 한계에 부딪혔다.

- 사물을 일반화하기 어려움: 예를 들어, 영상에서 사람을 인식할 경우 사람의 표준적인 특징을 어떻게 뽑아야 할까? 사람마다 신체를 구성하는 기본 요소가 있으므로 분명히 일반화할 여지가 있지만 알고리즘적으로 파악하기 위해서는 훨씬 더 구체적인 정의가 필요하다. 만일 다른 사물에 가려서 기본 요소 중 일부를 파악할 수 없을 경우 어떻게 할 것인가? 사람마다 피부색과 신장이 다른 데 이를 어떻게 처리할 것인가? 활발하게 움직일 경우 어떻게 대응할 것인가?
- 기존 사물의 특징이 바뀌거나 다른 사물을 추가할 때마다 소프트웨어 자체를 변경해야 하는 어려움: 예를 들어, 차량을 인식할 경우 차량 모델 별로 특징을 뽑아야 하는데, 차량 외형은 매년 조금씩 바뀌고 어떤 경우에는 완전히 새로운 차종이 등장하기도 한다. 이럴 때마다 소프트웨어의 알고리즘을 변경해야 하는데 이는 제조사 입장에서 큰 부담으로 다가온다. 복잡도가 점점 높아지는 설계와 구현 과정은 물론이고 소프트웨어 변경에 따르는 치밀한 검증과 오동작을 유발할 가능성을 줄이기 위한 완벽한 배포 과정도 생각해야 한다.



위와 같은 한계점으로 인해 컴퓨터가 영상에서 사물을 인식하는 능력은 사람에 미치지 못하는 상황이었다. 사람 수준에 도달하지 못한다면 적극적인 외부 도움 없이는 자율주행도 제한적일 수밖에 없다.

## 2. 학습이 가능한 컴퓨터

그렇다면 사람을 대신해 컴퓨터가 데이터를 받아들여 스스로 학습하면 어떨까? 매년 사물의 변화에 맞춰 알고리즘을 고안하고 이를 프로그램으로 옮기는 대신에 처음부터 학습이 가능한 일반적인 알고리즘을 만들어 프로그램으로 구현한 다음에 여러 가지 다양한 영상 데이터를 넣어 훈련을 해서 특정 사물을 파악하는 원리를 스스로 찾아내게 만들면 변화에 빠르게 대응할 수 있다. 전통적인 방식에서는 사람이 알고리즘을 만들고 이를 토대로 구현한 프로그램에 데이터를 입력해 처리하는 방법을 사용했다면, 이와 반대로 기계 학습은 학습이 가능한 프로그램을 만들고 데이터를 입력하여 알고리즘을 만들어내는 방법을 사용한다.

기계 학습의 핵심은 데이터를 평가하는 표현(representation)과 아직 알려져 있지 않은 데이터의 처리인 일반화(generalization)에 있다. 기계 학습은 미리 주어진 학습 데이터를 토대로 미래에 알려지지 않은 데이터가 들어올 경우 무엇인지 예측하는 능력을 제공하므로 전통적인 영상처리의 문제점 중 하나를 공략한다. 기계 학습은 통계적인 방법, 대수적인 방법, 기호적인 방법을 사용하여 구현하며, 경우에 따라서는 여러 가지 방법을 섞어서 문제를 해결하기도 한다. 하지만, 학습된 데이터를 이용하여 만들어진 모델에 근접한 새로운 데이터가 들어올 경우에는 정확하게 파악하는 반면 사람과 같은 융통성을 발휘해서 새롭지만 기존과는 살짝 다른 모호한 데이터를 처리하기 어렵다는 한계가 존재한다.

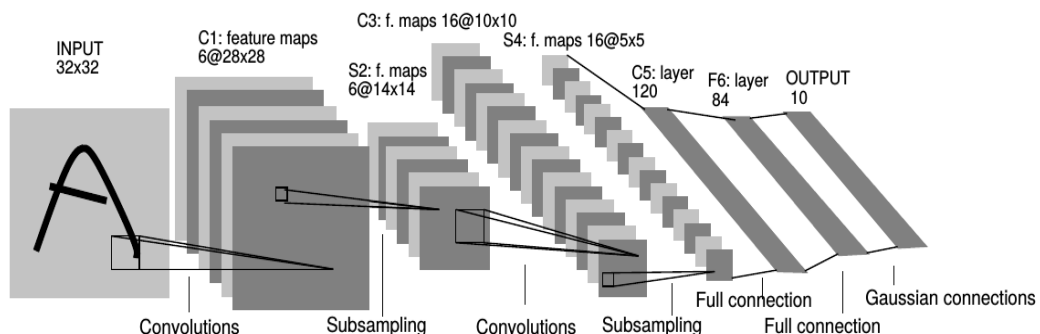
그렇다면 사람을 닮은 학습 방법을 컴퓨터에 구현할 수 있을까? 초창기 컴퓨터 기술이 처음 등장할 때부터 두뇌를 모사하는 방법에 대한 연구가 시작되었지만, 낮은 컴퓨터 성능과 학습에 필요한 데이터 부족으로 인해 답보 상태에 빠졌다. 긴 동면 끝에 역전파(back propagation) 기법을 재발견한 1980년대 중반 이후를 거쳐 경사 하강법(gradient descent) 방식의 한계점인 국소 최저값(local minima) 문제를 푸는 실용적인 전처리(preprocessing) 기법의 등장과 방대한 데이터를 처리할 수 있을 만큼 컴퓨터 성능이 높아짐으로써 2000년대 중반 이후에 딥러닝은 인공지능의 주류 기술로 떠올랐다.

### 3. 딥러닝과 영상 분석

딥러닝은 국소 최저값이라는 문제를 해결하기 위해 데이터의 전처리 과정을 도입해 비지도 학습법(unsupervised learning)으로 각 층을 손질한 다음에 처리한 데이터를 여러 층으로 쌓아 올려 최적화를 수행한다. 이와 같은 결과로 인해 여러 신경망 단계를 거쳐 복잡한 함수를 회귀(regression)/분류(classification)할 수 있게 됨으로써, 사람에 필적하는 융통성을 발휘하는 수준에 이르렀다.

딥러닝 역시 기계 학습의 범주이므로 최종 목표는 데이터를 사용하여 학습하고 학습 결과로 만들어진 모델을 사용하여 새로운 데이터의 특징(feature)을 예측하는 데 있다. 여기서 기계 학습 결과 모델이 예측을 제대로 해내기 위해 특징을 자동으로 학습하는 능력이 필요하다. 분류나 회귀 과정에서 사용하기 위해 자동으로 특징을 추출하려면 공통 패턴을 찾아내는 자동화된 알고리즘이 중요하다. 딥러닝에서 합성곱(convolutional) 층은 복잡도가 증가(모서리나 얼룩 → 눈/코/입 → 얼굴)하는 비선형 특징의 계층을 이루기 위해 이미지에서 찾아낸 좋은 특징을 다음 층으로 넘기는 과정에 적합하다. 마지막 층은 분류나 회귀를 위해 이 모든 일반화된 특징을 활용한다. 합성곱 신경망을 사용할 경우, 비선형 특징의 여러 층을 추출해 분류기에 넘겨 이 모든 특징을 합쳐 예측을 해낸다. 이 때 층이 너무 얇을 경우 복잡한 특징을 학습할 수 없으므로 아주 깊은 비선형 특징 계층을 차곡차곡 쌓아야 한다. 강력한 계산 능력을 제공하는 고성능 (GP)GPU와 수많은 학습 데이터가 필요한 이유가 바로 여기에 있다[4].

현재 가장 널리 쓰이는 영상 분석 기술인 CNN(Convolutional Neural Network)은 영상 분석을 위한 주류로 자리 잡고 있다. CNN은 유용한 정보를 얻기 위해 입력을 필터링하기 위해 합성곱



<자료> [5] Gradient-Based Learning Applied to Document Recognition

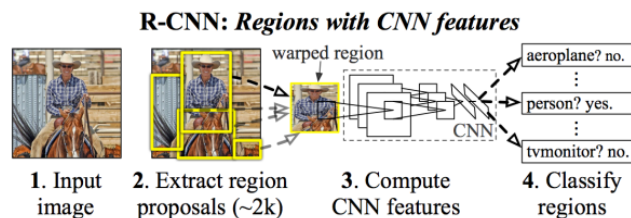
[그림 1] LeNet-5 아키텍처

계층을 사용하는데, 합성곱은 특징 맵(입력 데이터)과 합성 커널의 각 원을 곱한 다음 합계를 내어 새로운 특징 맵을 생성하는 수학적 연산이다. 합성곱 계층은 학습이 가능한 매개변수를 포함하므로 형태, 윤곽선, 색상과 같은 유용한 정보를 추출하게끔 학습 과정에서 합성 커널(필터)이 자동으로 조정된다. [그림 1]은 손글씨를 인식하기 위해 르쿤이 개발한 LeNet-5라는 CNN 참조 아키텍처를 보여준다. 여러 계층으로 나누어진 합성곱과 서브샘플링 과정을 거쳐 최종적으로 글씨를 파악하는 구조로 되어 있다.

#### 4. 객체 판별을 넘어서 검출로[6],[7]

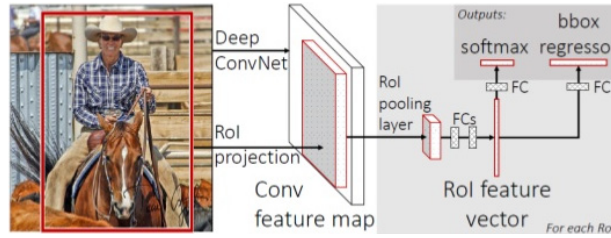
CNN은 영상이 담고 있는 특정 객체 하나가 무엇인지 판별해내는 능력에서 사람과 대등한 수준에 이르렀다. 하지만 자율주행차량이 수집하는 영상 정보는 특정 객체 하나가 아니라 여러 차량, 사람, 표지판 등 다양한 위치에 다양한 객체를 포함하고 있으므로 판별해야 할 대상 객체가 어떤 영역에 위치하는지부터 찾아야 한다. 게다가 영상에 포함된 객체들은 서로 중첩되거나 배경에 묻혀버리기도 한다. 전통적인 방법을 사용하여 일단 객체가 있을만한 위치를 파악한 다음에 판별 과정으로 넘어가도 되지만, 객체가 있을만한 위치를 파악하는 작업 자체가 이미 객체를 판별하는 과정을 어느 정도 포함하고 있기 때문에 검출 과정에서 효율성과 정확성에 제약이 오기 마련이다.

초기에는 R-CNN(Region with CNN)이라는 기법이 등장했다([그림 2] 참조). R-CNN에서는 전처리 과정으로 원하는 영상에 선택적 탐색을 적용하여 객체가 있으리라 추정하는 후보 영역(Region Proposal: RP)을 찾아낸 다음에 각 후보 영역마다 CNN을 적용하는 방법을 사용한다. 후처리 과정으로 전통적인 SVM(Support Vector Machine) 기법을 활용하여 CNN에서 얻은 결과물을 판별한다. 하지만 전반적인 검출 과정에서 RP마다 CNN을 적용해야 하며 RP 자체도 계산이 필요하므로 성능 문제가 발생한다.



<자료> [8] Rich feature hierarchies for accurate object detection and semantic segmentation

[그림 2] R-CNN 아키텍처

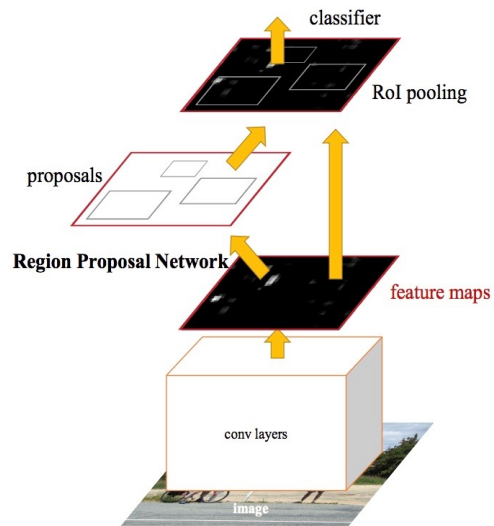


<자료> [9]Fast R-CNN

[그림 3] Fast R-CNN 아키텍처

이와 같은 R-CNN의 문제를 해소하기 위해 Fast R-CNN이라는 개선된 기법이 뒤를 이었다 ([그림 3] 참조). Fast R-CNN에서는 ROI(Region of Interest) 풀링(pooling, 계산량을 줄이고 과적합(overfitting)을 방지하기 위해 특정 영역에서 최대값이나 최소값을 뽑아내는 통합 과정)이라는 층을 도입하여 CNN의 특징 맵의 일부 영역에서 정규화된 특징을 추출한다. 따라서 입력 영상에 CNN을 한 번만 적용하고 ROI 풀링으로 객체 판별을 위한 특징을 추출한다. 따라서 RP 계산은 물론이고 RP마다 반복적인 CNN 적용도 필요하지 않으므로 성능을 높일 수 있다. 또한, R-CNN과는 달리 Fast R-CNN은 영상 판별을 위해 SVM을 사용하지 않고 단일 네트워크 내부에 소프트맥스(softmax, 입력 과정에서 받은 값을 정규화시켜 0과 1사이의 출력값을 산출해내는 함수이며, 모든 출력값의 총합은 1이 되는 특성이 존재한다) 층을 배치하는 방법으로 모델 단순화할 수 있다.

하지만 Fast R-CNN에도 병목 지점이 존재한다. 바로 후보 영역(RP)을 찾는 부분이며, 선택적 탐색 비용이 비싸기 때문에 이를 해결해야 성능을 개선할 수 있다. Faster R-CNN은 독자적인 선택적 탐색 알고리즘을 대신하여 후보 영역 제안을 위해 동일한 CNN 결과를 재사용하는 현명한 방식을 택한다([그림 4] 참조). 즉, 후보 영역 검출과 객체 판별 목적으로 CNN 하나만 훈련하면 된다.



<자료> [12] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

[그림 4] Faster R-CNN 아키텍처

## 5. 데이터, 학습, 시뮬레이션

제대로 된 학습 모델을 생성하려면 학습이 가능하도록 만들어진 양질의 풍부한 데이터가 필요하다. 다양한 센서로부터 수집된 데이터를 정제한 다음에 객체 위치를 표현하는 바운딩 박스와 함께 객체 유형과 의미를 부여하는 레이블링 작업을 거쳐 딥러닝 기법으로 학습을 진행한다.

이처럼 학습에 필요한 데이터가 점점 더 중요해짐에 따라 자율주행차량 제조사들은 다양한 방법으로 데이터를 수집하고 있다. 웨이모나 우버의 경우 독자적으로 카메라와 라이다가 장착된 차량을 운행하면서 고해상도 데이터를 수집하고 있으며, 자율주행차량을 위한 영상처리 부품 업체로 유명한 모빌아이의 경우 REM(Road Experience Management)[14]라는 프로그램을 운영하여 대중들로부터 데이터를 수집하기도 한다. 대학교를 비롯한 연구기관에서도 자율주행차량 학습을 위한 여러 데이터를 제공하고 있다. 캘리포니아 공과대학교에서는 도심 환경의 일반 주행 과정에서 얻은 비디오 영상을 가공해 보행자 레이블링 데이터를 제공하고 있으며 [11], 버클리 공과대학에서는 <Berkeley DeepDrive>[15]라는 사이트를 만들어 다양한 기상, 시각, 상황을 반영한 도심 운행 영상 데이터와 도로 위에 존재하는 객체 레이블링 데이터를 제공하고 있다.

앞서 소개한 CNN 계열의 학습 알고리즘을 사용하여 방대한 데이터를 모델로 변환하는 작업을 거치게 된다. 학습 시간을 단축하기 위해 대규모 GPU 클러스터를 활용하여 병렬로 학습을 수행하게 되며, 이렇게 만들어진 모델을 시뮬레이터에서 돌려 강화학습을 진행하면 점점 더 개선된 결과를 얻게 된다. 웨이모는 일찌감치 카크래프트라는 가상현실을 접목한 시뮬레이터로 자율주행차량의 모델을 개선하는 작업을 벌이고 있으며[12], 바이두는 자율주행차량 개발에 도움이 되는 아폴로[16]라는 시뮬레이터를 마이크로소프트 애저(Azure) 클라우드 상에서 서비스 형태로 제공하고 있으며, 트루비전[17]은 자율주행을 위한 시뮬레이터를 오픈소스로 공개했다.

## IV. 결론 및 시사점

미 국방성 소속의 방위고등연구계획국(Defense Advanced Research Projects Agency: DARPA)이 정부차원에서 자율주행차량의 연구개발 지원을 위해 2004년 처음으로 그랜드챌린지 대회를 열었을 때 완주한 차량은 하나도 없었다[13]. 그나마 카네기멜론 대학교팀이 240km 구간

중에서 12km에 못 미치는 기록을 세웠을 뿐이었다. 하지만 여기서부터 출발한 대학교와 연구소 중심의 자율주행차량 기술에 대한 연구 개발은 2018년 현재 테슬라, GM, 포드, 벤츠, 볼보를 비롯한 전세계 주요 완성차 업체는 물론이고, 웨이모와 바이두와 같은 기술 서비스 업체, 우버와 리프트와 같은 차량 공유 서비스 업체 전반으로 확산되었으며, 고성능 하드웨어는 물론이고 빅데이터 기술과 인공지능 기술의 발전과 함께 일부 지역에서는 거의 상용화가 가능한 수준까지 도달했다.

자동차 대국인 미국의 경우에는 자율주행차량에 법적인 지위를 보장하기 위해 면허 시스템부터 개선하고 있으며, 자율주행차량이 운행될 경우 발생할 수 있는 윤리적이고 도덕적인 문제(예: 트롤리 딜레마 - 인명사고를 피할 수 없는 상황에서 누구의 생명을 우선적으로 구할 것인가?)부터 시작해서 경제적인 문제인 손해배상 책임에 이르기까지 전방위에 걸쳐 전통적인 시스템의 부족한 부분을 채우기 위해 노력하고 있다.

자율주행차량은 숙박, 주차, 보험, 물류, 대중교통 등 주변 산업에 큰 영향을 미치므로 원천 기술을 사전에 확보해야 국가 경쟁력을 유지할 수 있다. 자율주행차량의 핵심인 센서와 CPU/(GP)GPU로 대표되는 하드웨어는 물론이고 인공지능 학습과 예측이 가능하게 만드는 소프트웨어의 기술 수준을 함께 발전시켜야 경쟁력을 확보할 수 있다. 특히, 인공지능 기술 중에서도 영상처리에 강한 딥러닝 기술에 신경을 써야 하며, 정교한 모델을 만들기 위해 양질의 대규모 데이터를 확보해야 할 것이다. 이와 동시에 자율주행차량이 상용화될 경우에 대비하여 법적인 준비도 갖춰야 하며, 공유차량 서비스와 같은 부가가치 서비스가 가능하도록 각종 규제에 대한 전반적인 검토도 필요하다. 정부 차원에서 자율주행차량에 대한 집중적인 관심과 연구 지원 및 투자를 기대한다.

#### [ 참고문헌 ]

- [1] Dmitri Dolgov, "Google I/O Recap: Turning self-driving cars from science fiction into reality with the help of AI," Google, 2018. 5. 9.
- [2] Stefano Markidis, Steven Wei Der Chien, Erwin Laure, Ivy Bo Peng, Jeffrey S. Vetter, "NVIDIA Tensor Core Programmability, Performance & Precision," arXiv:1803.04014 [cs.DC], 11[v1], 2018. 5, p.8.
- [3] Norman P. Jouppi, "In-Datacenter Performance Analysis of a Tensor Processing Unit," et al, arXiv:1704.04760 [cs.AR], 16[v1], 2017. 4, p.11.
- [4] Tim Dettmers, "Deep Learning in a Nutshell: Core Concepts," nVidia, 2015. 11. 3.
- [5] Yann Lecun, "Gradient-Based Learning Applied to Document Recognition," Proc. of the IEEE, Nov. 1998, p.7.

- [6] Dhruv Parthasarathy, "A Brief History of CNNs in Image Segmentation: From R-CNN to Mask R-CNN," Athelas, 2017. 4. 22.
- [7] 카카오시리포트, "사진에서 사람을 읽다", 카카오 정책산업 연구, 2017. 5. 19.
- [8] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," arXiv:1311.2524 [cs.CV], 2013. 11. 11[v1], 2014. 10. 22[v5], p.1.
- [9] Fast R-CNN, Ross Girshick, arXiv:1504.08083 [cs.CV], 2015. 4. 30[v1], 2015. 10. 27[v2], p.2.
- [10] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," arXiv:1506.01497 [cs.CV], 2015. 6. 4[v1], 2016. 1. 6.[v3], p.3.
- [11] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," PAMI, 2012, p.2.
- [12] Cade Metz, "What Virtual Reality Can Teach a Driverless Car," The New York Times, 2017. 10. 29.
- [13] Evan Ackerman, "Carnegie Mellon Solves 12-Year-Old DARPA Grand Challenge Mystery," IEEE Spectrum, 19 Oct. 2017, pp.1-2.
- [14] <https://www.mobileye.com/our-technology/rem/>
- [15] <http://bdd-data.berkeley.edu/>
- [16] <http://apollo.auto/platform/simulation.html>
- [17] <https://www.truevision.ai/>