

패턴인식 보고서

120220210 고재현

2022년 12월 19일

1 개요

1.1 목적

- 최근 출시되는 자율주행 자동차에 포함된 pattern recognition/computer vision 관련 요소기술을 조사한다.
- 해당 요소기술과 수업에서 다룬 주제들 간의 연관성을 파악한다.

1.2 선정 모델 및 브랜드

선정한 모델은 **Tesla Model S** 이다. 테슬라가 오토파일럿 기능을 앞세워 자율주행 시장을 선도하고 있기 때문이다. 해당 모델에 적용된 pattern recognition/computer vision 관련 요소기술은 다음과 같다.

- **Autopilot**: 자율주행 기능
- **Autopark**: 주차 자동화 기능
- **Autosteer**: 자동 조향 기능

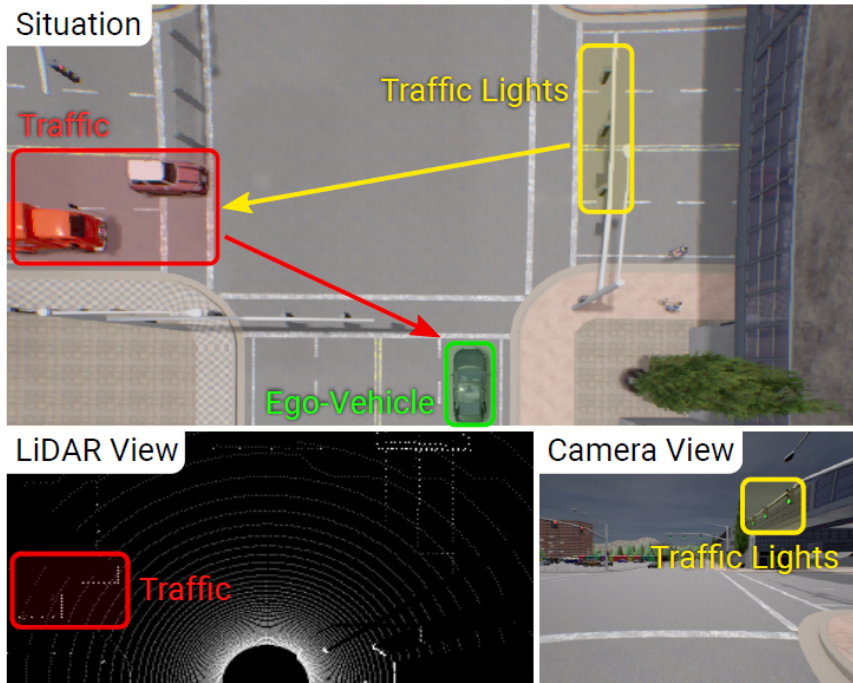
1.3 선정 논문

[?, Multi-modal fusion transformer for end-to-end autonomous driving] 및 [?, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors] 논문을 선정하였다. 첫 번째 논문을 선정하게 된 이유는 자율주행 시스템의 구성요소 중 하나인 perception에 대한 연구이면서, 필자의 연구 분야인 multi-modal fusion에 대한 연구이기 때문이다. 두 번째 논문을 선정하게 된 이유는 자율주행 시스템의 구성요소 중 하나인 perception에 대한 연구이면서, 영상처리의 주요한 분야 중 하나인 object detection에 대한 연구이기 때문이다.

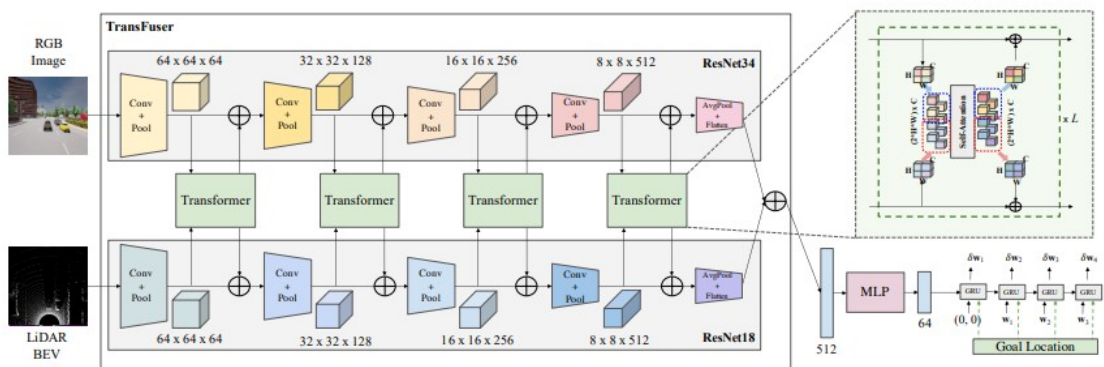
2 논문 요약

2.1 Multi-modal fusion transformer for end-to-end autonomous driving

이 논문은 그림 1의 상황처럼 라이더(LiDAR : Light Detection And Ranging) 센서로 얻을 수 있는 주변의 차량의 위치에 따른 교통정보와 카메라로 얻을 수 있는 신호기에 따른 교통정보가 다른 경우, 두 센서로부터 얻을 수 있는 정보를 결합하여 차량의 주행을 제어하는 것을 목적으로 한다. 그림 2는



〈그림 1〉 논문에서 해결하려는 문제 상황



〈그림 2〉 Transfuser 구조

Transfuser의 구조를 보여준다. 두 센서의 출력으로부터 Resnet 구조[?]를 이용하여 정보를 추출하는 과정에서, 각 layer의 출력단으로부터 추출된 정보를 Transformer[?]를 이용하여 결합하는 것을 확인할 수 있다.

2.1.1 method

목적 논문에서 제시한 목표는 시내 도로 주행에서의 point-to-point navigation이다. point-to-point navigation은 차량이 목표지점까지 waypoint를 따라 교통법규를 지키면서 다른 차량과의 상호작용을 하며 완주하는 것을 의미한다. 이를 달성하기 위한 방법으로 강화학습 기법 중 하나인 Imitation Learning을 채용하였다. Imitation Learning은 전문가가 직접 주행한 데이터를 따라하도록 agent의 policy를 학습하는 것을 의미한다.

입출력 자율주행 오픈소스 시뮬레이터 CARLA[?]에 있는 urban 가상환경에서 수집한 데이터를 입출력으로 사용하였다. 입력은 두 가지로, 카메라와 라이다 센서의 출력이다. 카메라로부터 얻은 영상의 왜곡을 줄이기 위해 이미지 입력의 중앙을 잘라내어 $256 \times 256 \times 3$ 크기로 사용했다. LiDAR 센서의 출력 또한 주변부분의 정보를 기반으로 $256 \times 256 \times 2$ 사이즈로 잘라내어 사용하였다. 채널의 한쪽은 지면 위, 한쪽은 지면 아래를 의미한다. 출력은 PID controller로 차량을 제어하기 위해 4개의 waypoint $\{w_t = (x_t, y_t)\}_{t=1}^T$ 로 설정했다.

모델 모델은 그림 2에서 두 가지 부분으로 나눌 수 있다. 첫째는 Resnet과 Transformer를 이용하여 구성한 Multi-Modal Fusion Transformer(Transfuser)이고, 둘째는 MLP와 GRU로 구성된 Waypoint Prediction Network이다. 먼저 Transfuser의 동작을 살펴보자. 전반적인 동작은 subsection 2.1의 표제 문단에서 작성하였으므로 Transformer의 적용 방법만을 확인한다. Transformer로는 GPT 모델을 사용하였다.

- 라이다 입력과 영상 입력에 대해 컨볼루션과 풀링을 진행하여 채널 수를 늘리면서 특징을 추출한다.
- 특징의 크기를 average pooling을 통해 8×8 로 줄인다.
- 각 특성맵을 concat하여 16×8 크기의 특성맵을 만든다.
- velocity를 value로, 16×8 특징을 key와 query로 사용하여 self attention(dot product attention)을 적용한다.
- bilinear interpolation을 통하여 원본 영상의 크기로 확대한다.
- 이전 단의 특성맵과 attention을 통해 추출한 특성맵을 더하여 특성맵을 업데이트한다.

둘째로 Waypoint Prediction Network의 동작을 살펴보자.

- activation이 ReLU인 3-Layer Perceptron을 이용하여 $1 \times 1 \times 512$ 크기의 특징을 $1 \times 1 \times 64$ 크기의 특징으로 압축한다.
- 압축된 특징을 GRU에 입력하여 4개의 waypoint를 예측한다.

Method	Town05 Short		Town05 Long	
	DS \uparrow	RC \uparrow	DS \uparrow	RC \uparrow
CILRS [16]	7.47 ± 2.51	13.40 ± 1.09	3.68 ± 2.16	7.19 ± 2.95
LBC [8]	30.97 ± 4.17	55.01 ± 5.14	7.05 ± 2.13	32.09 ± 7.40
AIM	49.00 ± 6.83	81.07 ± 15.59	26.50 ± 4.82	60.66 ± 7.66
Late Fusion	51.56 ± 5.24	83.66 ± 11.04	31.30 ± 5.53	68.05 ± 5.39
Geometric Fusion	54.32 ± 4.85	86.91 ± 10.85	25.30 ± 4.08	69.17 ± 11.07
TransFuser (Ours)	54.52 ± 4.29	78.41 ± 3.75	33.15 ± 4.04	56.36 ± 7.14
<i>Expert</i>	84.67 ± 6.21	98.59 ± 2.17	38.60 ± 4.00	77.47 ± 1.86

〈그림 3〉 Transfuser 실험 결과

2.1.2 result

2.2 YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

asdf