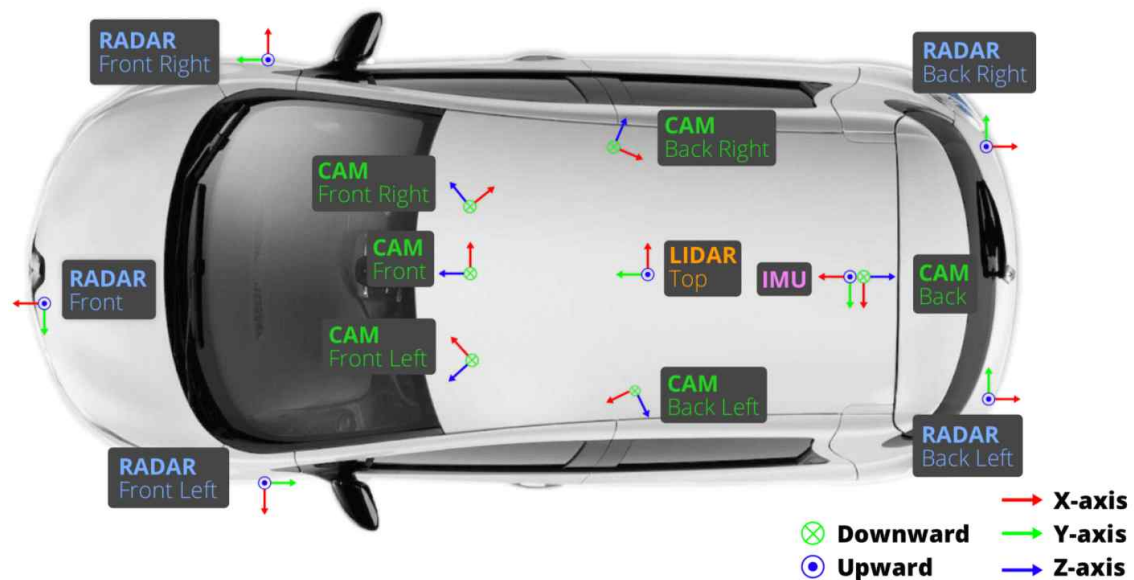


1. 자율주행 인지에 관련된 3종 이상의 공개 Data Set 조사, 정리
  - 데이터 설명, 양, 세부요소(feature), 데이터 예시, 활용 예 등

#### 1) nuScenes Data Set

nuScenes Data Set는 Motional(이전의 nuTonomy)팀이 개발 한 자율주행을 위한 공개 대규모 Data Set이다. Motional은 무인 차량을 안전하고 신뢰할 수 있으며 접근 가능한 현실로 만드는 것을 목표로 한다. 이를 위해 교통량이 많고 운전 상황이 매우 까다로운 두 도시인 보스턴과 싱가포르에서 1000개의 운전 장면을 수집한다. 20초 길이의 장면은 다양하고 흥미로운 운전, 교통 상황 및 예상치 못한 행동을 담고 있다. nuScenes Data Set의 경우 보스턴과 싱가포르에서 약 15시간의 운전을 통해 Boston Seaport 및 싱가포르의 One North, Queenstown 및 Holland Village에서 데이터를 수집한다. 다양한 시나리오를 포착하기 위해 다양한 위치, 시간 및 기상 조건을 목표로 하고, 클래스 빈도 분포의 균형을 맞추기 위해 희귀 클래스(예:자전거)가 있는 장면을 더 많이 포함하여 운전 경로가 신중하게 선택된다.

nuScenes의 풍부한 복잡성은 장면 당 수십 개의 물체가 있는 도시 지역에서 안전한 운전을 가능하게 하는 방법의 개발을 타겟으로 한다. 다른 대륙에서 데이터를 수집하면 다양한 위치, 기상 조건, 차량 유형, 초목, 도로 표시 및 왼손 대 오른손 교통 전반에 걸친 컴퓨터 비전 알고리즘의 일반화를 연구할 수 있다.

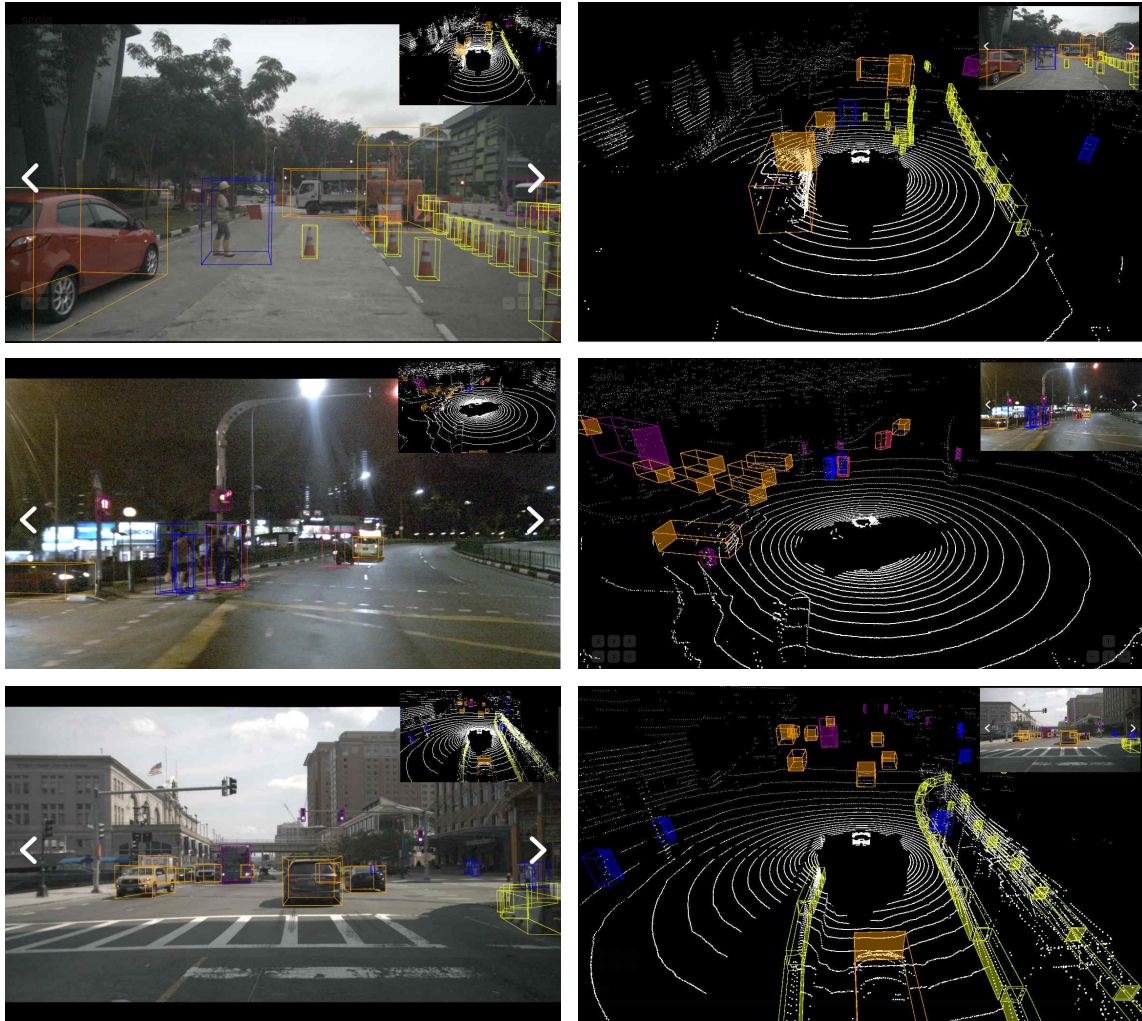


< Car setup - Renault Zoe>

차량의 세팅은 1개의 spinning LIDAR, 5개의 long range RADAR sensor, 6개의 camera, GPS, IMU로 이루어져 있다. 전체 데이터 세트에는 약 140만 개의 Camera image, Radar sweeps, 39만개의 LiDAR sweeps, 140만개의 object bounding box로 구성되어 있다. 추가 기능(지도 레이어, 원시 센서 데이터 등)이 곧 추가될 예정이다. nuScenes 데이터 세트는 KITTI 데이터 세트에서 영감을 받았지만, KITTI에 비해 nuScenes에는 7배 더 많은 객체 주석이 포함되어 있다. Object detection, tracking과 같은 Computer Vision 과제에 이용하기 위하여, 전체 dataset에 3D Bounding box와 함께 23개의 object class의 라벨링을 진행한

다. 또한 가시성, 활동 및 포즈와 같은 객체 수준 속성에 주석을 추가하게 된다.

BDD/Cityscapes/ApolloScapes와 같은 카메라 기반의 data set과는 차별화되는 점은 nuScenes는 모든 센서를 다루는 것을 목표로 한다는 것이다.



< 다양한 주행 환경(낮, 밤, 흐린 날씨 등)에서의 데이터 예시(Scenes) >

첫 번째 nuScenes 릴리스에서는 경계 상자 또는 입방체가 3D 개체를 나타내는 데 사용되고 많은 경우에 유용하지만 직육면체는 관절이 있는 개체의 미세한 모양 세부 사항을 캡처하는 기능이 부족하다. lidar semantic segmentation의 약자인 nuScenes-lidarseg는 의미론 레이블이 있는 nuScenes 데이터 세트의 40,000개 키 프레임에 있는 모든 단일 라이다 포인트에 대한 주석을 포함하여 더 높은 수준의 세분성을 제공한다. nuScenes의 23개의 전경 클래스 외에도 9개의 배경 클래스를 포함하였다.

nuScenes-lidarseg의 분류는 나머지 nuScenes 및 nuImages와 호환되므로 여러 센서 양식에 대한 광범위한 연구가 가능하다. 이는 연구자들이 점 수준 의미론을 사용하여 라이다 포인트 클라우드 세분화, 전경 추출, 센서 보정 및 mapping과 같은 새로운 문제를 연구하고 정량화할 수 있도록 활용 분야를 넓힌다.

### ① Lidar segmentation & Prediction

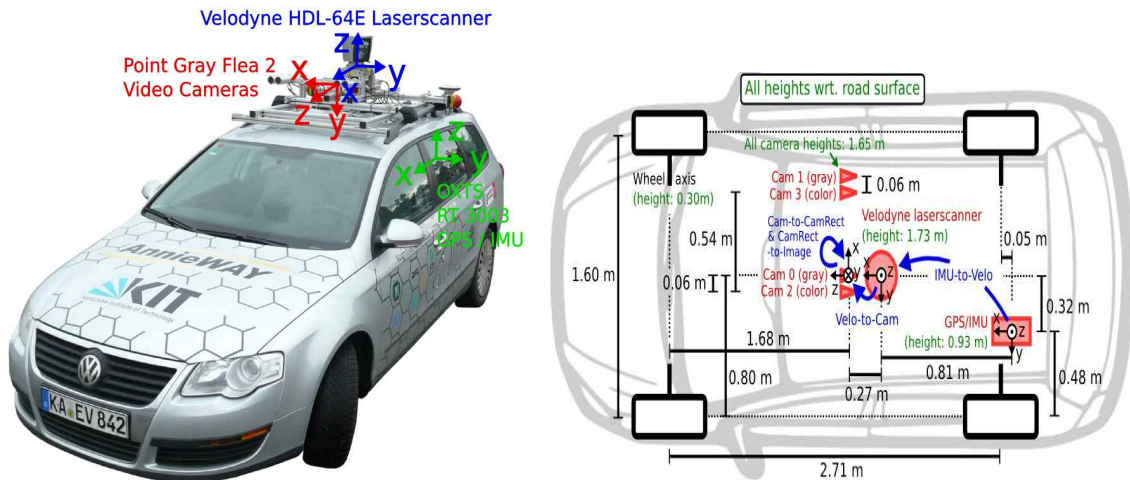
nuScenes data set은 nuScenes-lidarseg를 이용하여 segmentation 과제를 수행할 수 있다. 과제 목표는 point clouds set의 모든 point에 대하여 카테고리를 prediction 하는 것이다. Task의 평가 metrics로는 mean intersection-over-union(mIOU)를 사용한다. 또한 nuScenes data set을 이용하여 object의 경로를 예측할 수도 있다. Task의 평가 metrics로는 세 가지를 사용한다. 예측된 경로와 ground\_truth 상의 point들 사이의 L2(유클리디안) 거리를 평균 내는 'minADE\_k', 예측된 경로와 ground\_truth의 최종 목표 지점 사이의 거리를 이용하는 'minFDE\_k', 예측된 경로와 ground\_truth의 L2 거리의 최댓값이 2m보다 크다면 예측에 실패했다고 정의하는 세 가지 방식을 사용한다.

### ② Detection & tracking

nuScenes data set을 활용한 3D object detection task의 목표는 각 set의 특성 및 속도 vector의 추정뿐만 아니라 서로 다른 10개의 카테고리에 3D bounding box를 만드는 것이다. nuScenes의 23개의 class 중, 비슷한 것은 서로 군집화하고 유사성이 없는 없는 class는 생략함으로써 10개의 클래스를 이용한다. 예를 들어, 오토바이/트럭/버스/자동차는 vehicle이라는 하나의 클래스로 묶는 것이다. Detection 결과들은 2Hz key frame 단위로 평가되며 train/validation/test set 모두 json 형태로 저장된다. Task의 평가 metrics로는 mAP, TP 성능 지표를 활용한다. Tracking은 detection에서 자연스레 이어지는 과정이다. 잘 알려진 detection 알고리즘으로부터 얻어진 object를 시간이 지남에 따라 추적하는 것이다. camera, lidar, radar 센서를 이용하여 3D multi object tracking을 진행하고, online tracking을 실시한다.

## 2) KITTI Data Set

KITTI(Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago)는 현재 자율주행에 사용되는 대표적인 Data Set중 하나다. 제공되는 Data Set은 stereo, optical flow, visual odometry, 2/3D object detection 및 3D tracking과 같은 다 분야의 컴퓨터 비전 연구 분야를 위해 만들어졌다. 이를 위해 고해상도 color/grayscale 비디오카메라 2개가 장착된 표준 스테이션 왜건을 장착했고, Velodyne사의 레이저 스캐너와 GPS 위치 확인 시스템을 통해 정확한 Ground Truth가 제공된다. 독일의 중형 도시인 Karlsruhe 주변 시골 지역 및 고속도로와 같은 환경에서 Annieway를 주행하며 제작되었다. 이미지 당 최대 15대의 자동차와 30명의 보행자를 확인할 수 있으며, 모든 데이터를 원시 형식으로 제공하는 것 외에도 각 작업에 대한 bench mark를 추출한다. 각 bench mark에 대해 평가 지표와 평가 웹 사이트도 제공한다.



### < Car setup - Annieway >

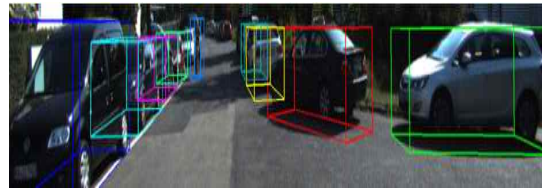
차량의 세팅은 1 Inertial Navigation System(GPS/IMU), 1 Laser scanner(Velodyne), 2 Grayscale cameras, 2 Color cameras, 4 Varifocal lenses로 구성 되어 있다. 레이저 스캐너는 초당 10프레임으로 회전하여 주기 당 약 100k 포인트를 캡처한다. 레이저 스캐너의 수직 해상도는 64이며 카메라는 접지면과 거의 수평으로 장착된다. 카메라 이미지는 libdc의 형식 7모드를 사용하여 1382 x 512 픽셀 크기로 잘리며, 수정 후 이미지가 약간 작아진다. 카메라는 동적으로 조정된 셔터 시간 (최대 셔터 시간 : 2ms)으로 레이저 스캐너 (앞을 향할 때)에 의해 초당 10 프레임으로 trigger된다.

raw 데이터는 6개의 카테고리(도시, 주거, 도로, 캠퍼스, 사람, calibration)로 이루어져 있으며, 약 13000개의 이미지로 구성된다. stereo, flow, secenflow의 경우 각각 약 200개의 training scenes와 test scenes으로 구성된다. depth의 경우 원시 LiDAR 스캔 및 RGB 이미지와 함께 93000개 이상의 depth map이 포함되어 있다. Visual Odometry의 경우에는 손실이 적은 png 형식으로 저장된 22개의 스테레오 시퀀스로 구성된다. 2/3D object detection에는 7481개의 training scenes와 7518개의 test scenes로 구성된다. 도로 및 차선 추정 벤치마크는 289개의 raining scenes, 290개의 test scenes로 구성된다.

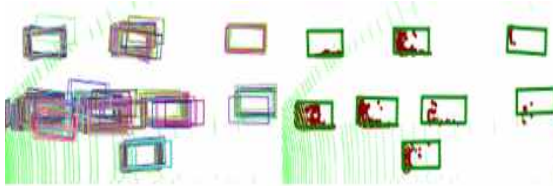




< MOTS trainset image >



< 3D object detection >



< Bird's Eye View >



< semantic segmentation >



< Visual odometry >



< Raw data >

① 2/3D Object Detection : 2/3D object detection에는 7481개의 training scenes와 7518개의 test scenes로 구성된다.

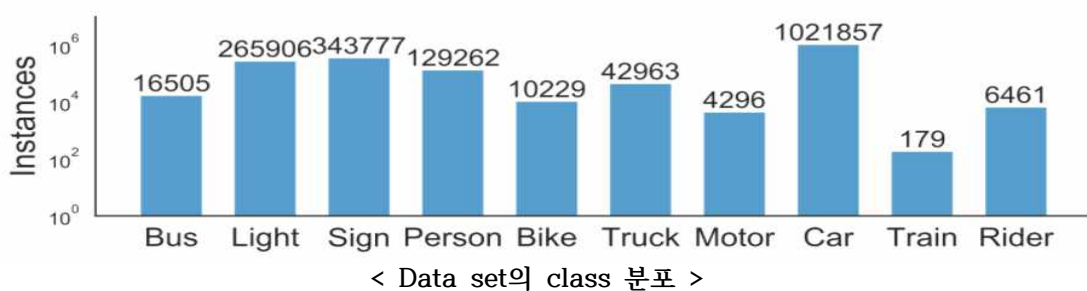
② Scene Flow : 약 200개의 training scenes와 test scenes으로 구성된다.

③ Visual Odometry : Visual Odometry benchmark는 연속된 22개의 stereo로 구성되어 있으며 이 중 11개는 training을 위해, 나머지는 evaluation을 위해 구성되어 있다.

### 3) BDD-100k Data Set

UC 버클리 인공 지능 연구 실험실(BAIR)에서 BDD(Berkeley Deep Drive)-100K로 불리는 운전 데이터 베이스를 공개했다. 자율주행을 더 안전하게 만들기 위한 인식 알고리즘의 경계를 탐구하는 데 관심이 있기 때문에 잠재적인 알고리즘을 설계하고 테스트하기 위해 실제 운전 플랫폼에서 수집한 데이터의 모든 정보를 활용하고자 한다. 데이터에는 4가지 주요 속성이 있다. 대규모이고 다양하며 거리에서 캡처되고 시간 정보가 포함된다. 데이터 다양성은 인식 알고리즘의 견고성을 테스트하는데 특히 중요하지만, 현재 열려있는 Data Set는 위에서 설명한 속성의 하위 집합만 포함할 수 있다. 따라서 Nexar의 도움으로 지금까지 컴퓨터 비전 연구를 위한 가장 크고 다양한 개방형 주행 비디오 Data Set인 BDD-100K 데이터베이스를 출시하였다.

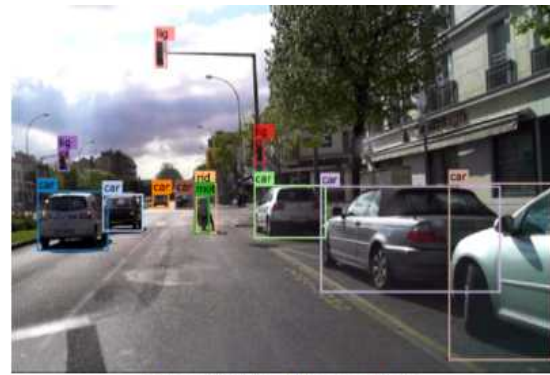
이름에서 알 수 있듯이 Data Set은 10만개의 동영상으로 구성되며 각 비디오의 길이는 약 40초, 720픽셀 해상도 및 30fps이다. 이 동영상들은 7만 개의 training set, 1만 개의 validation set, 2만 개의 testing set으로 구성되었고, 각 동영상에서 10초에 key frame을 sampling하고 해당 key frame에 대한 주석을 제공한다. image tagging, road object bounding boxes, drivable areas, lane markings, and full-frame instance segmentation과 같은 여러 level로 레이블이 지정된다. 동영상에는 거친 주행 환경 구현, GPS 정보, IMU 데이터 및 타임 스탬프가 포함되어 있다. 녹화된 비디오는 5만회의 운행을 통해 비오는 날씨, 흐린 날씨, 맑은 날씨, 안개와 같은 다양한 날씨 조건이 기록되어 있다. Data Set는 낮과 밤이 적절한 비율로 기록되어 있다. 이미지에 쉽게 주석 처리하기 위해, 수직 차선은 적색, 평행 차선은 청색으로 구분하였다. 적색 표시 있는 운전 경로와 청색 표시가 있는 대안 운전 경로로 주행 가능 구역을 구분한다. Data Set은 버스, 신호등, 교통 표지, 사람, 자전거, 트럭, 모터, 자동차, 기차 및 라이더를 위해 100,000개 이미지에 주석이 달린 2D Bounding Box가 포함되어 있다.



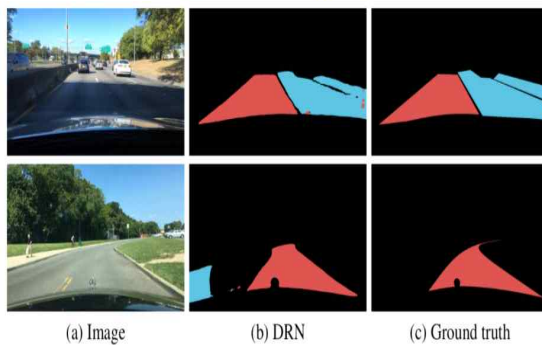
차선 표시는 운전자에게 중요한 도로 지침이다. 또한 GPS, 지도에 정확한 global coverage가 없을 때 자율주행 시스템의 주행 방향 및 위치 결정에 중요한 단서이다. 차선에서 차량을 지시하는 방법에 따라 차선 표시를 두 가지 유형으로 나눈다. 수직 차선 표시 (아래 그림에서 빨간색으로 표시)는 차선의 주행 방향을 따르는 표시를 나타낸다. 평행 차선 표시 (아래 그림에서 파란색으로 표시)는 차선에 있는 차량이 정지해야 하는 것을 나타낸다. 또한 실선 대 파선 및 이중 대 단일과 같은 표시에 대한 속성을 제공한다.



Lane Marking



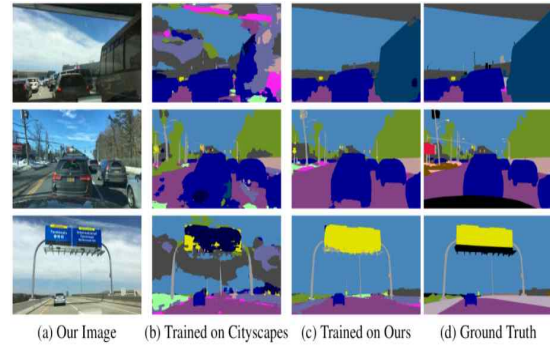
Road Object Detection



(a) Image

(b) DRN

(c) Ground truth



(a) Our Image

(b) Trained on Cityscapes

(c) Trained on Ours

(d) Ground Truth

Labeling System으로 bounding box나 region annotation같은 다양한 종류의 주석작업을 수행한다. 이 데이터는 도로/포장도로의 보행자 탐지를 위해 사용할 수도 있다. 이를 위해, 현재 데이터 세트에는 85,000개가 넘는 보행자 인스턴스가 있다. 이 데이터베이스는 차량용 컴퓨터 비전 및 기계 학습을 연구하기 위해, 현재 백만 대의 자동차, 300,000개가 넘는 도로 표지판, 130,000명의 보행자 등으로 구성된 Berkeley Deep Drive (BDD) Industry Consortium의 지원을 받고 있다. BDD-100K는 거리에서 보행자를 감지하고 피하는 컴퓨터 비전 알고리즘 구현에 유용하다. 이 데이터 세트로 도로 객체 감지, 인스턴스 세그멘테이션, 운전 가능 지역, 차선 구분, 시멘틱 세그멘테이션과 같은 작업이 가능하다.

① Driveable Area : 도로 위의 차선이 불분명한 경우가 존재하는데 이럴 때는 차선만으로 차량이 주행이 가능하지 판단을 내리기에는 적합하지 않다. 이런 경우를 대비하여 운전 가능 지역을 directly driveable area와 alternatively drive area로 구분 짓고 각각 적색과 청색으로 표시하게 된다.

② Road Object Detection : BDD-100K data set에는 버스, 신호등, 교통 표지판, 자전거, 트럭, 오토바이, 자동차, 기차, 사람, 라이더를 위한 10만 개의 key frame에 2D Bounding Box와 라벨이 주석처리 되어있다. 이전의 다른 dataset에 비해 보행자에 대한 정보가 많기 때문에 보행자 검출 및 회피를 위하여 BDD100K를 활용할 수 있다.

③ Lane Marking : 차량의 운전 방향을 의미하는 수직 차선은 적색으로 표시하였고, 차량이 멈춰야하는 차선을 의미하는 수평 차선은 청색으로 표시한다.