

## Abstract

중요한 핵심 자료는 주행에 대해 얼마나 잘 알고 있는지이다. HDD(Honda Research Institute Driving Dataset)는 실제 환경에서 운전자의 행동에 대해 연구에 도움될 수 있는 데이터 셋이다. 데이터 셋은 각 다른 센서를 이용하면서 104시간들의 샌프란시스코 bay에서의 실제 운전자의 화면을 보여준다.

## 1. Introduction

운전에는 여러 상황을 판단하고 이해하는 수준이 필요하다.( 사람을 추적, 인지, 분별 상황 감지 etc.) 시각에 대한 인지에 대한 실적은 cnn과 같은 넘쳐나는 데이터셋으로 인해 급상승되었다. 그러나 무인 자동차의 기술을 달성하기 위해서는 더 좋은 이해도가 필요하다.

[7, 12, 22, 28, 4, 34, 23] 은 알고리즘적인 구성들의 핵심을 이끌어냈고 6,7,4 의 대용량 데이터로 object detection, semantic segmentation에 기여했다. 추가적으로 22의 데이터셋은 다른 날씨와 밝기 여건에 mapping과 robust localization 에 대한 시도를 보여주었다. 그러나, 이 데이터셋들은 더 높은 수준의 시도를 하지않았다. 우리는 semantic categories 에 일치하는 보행자들과 지나가는 장면들을 인지, 감지하는 것이 첫 단계이다. 이러한 장면을 이해하는데 앞서, 운전자의 행동과 교통 상황과의 상호작용 이해가 필요로하다.

이 이상적인 목표에 도달하기위해, 우리는 어떻게 운전자들이 행동하는지 보행자들과의 상호작용을 목적으로 HDD를 만들어냈다. 104시간의 샌프란시스코, 137의 세션, 자세한 사항은 세션3에서.

우리는 운전자의 행위들을 미리 정의시켜 상황분류 한 것을 Figure 1에서 설명한다. 노란 사선은 차의 위치를 GPS로 표현하였다. 4 레이어는 각 운전자의 행동들을 묘사한다. 첫 번째는 GOAL-ORIENTED 초록색으로 left turn(left upper의 사진)이다. 두 번째는 stimulus-driven action 빨간색으로 stop(Lower left의 사진)이다. 정체는 Cause 에 속한다. Cause의 멈추는 원인은 빨간색 레드박스로 표시된다. Attention은 보라색으로 보행자들을 표시한다. Figure 1에서는 또한 정체 때 멈춤이 아닌 left turn으로 보행자가 건너고 있는 동안 U turn을 하는 경향이 있다는 것을 두가지 다른 시나리오로 설명해준다.

이와 같은 것들과 함께, 우리의 데이터셋은 무인자동차의 도전적인 시도를 가능하게한다. 균등하지 않은 곳에서 측정, 운전자들과 보행자들의 상호작용, 다수의 처리, 섞여있는 상황, 위 논문에서는, 운전자 시점에서 운전자 행동에 묘사하지만, 결정을 내리거나 계획하거나 보행자들을 예측과 같은 다른 측면들을 내포한다.

## 2. Related work

우리는 현 데이터셋과 과거 데이터셋들과의 관계를 살펴보자.

### -Driving Scene Datasets.

KITTI [7] 은 카메라들을 포함한 적절한 센서들을 제공, 그들은 이 분야들에서 알고리즘개발을 위해 다른 기준들에서 실행되었다. [22]는 OXFORD에서의 날씨와 빛 조건들에 대한 데이터는 대표한다.

우리 데이터셋은 이에 보완적이다. 우리는 다양한 상황들에 운전자들의 행동들에 집중되었기 때문이다.

최근 [3,12,28,34,13]은 인기를 끌고 있다. [3]은 TORCS 운전게임에 대한 적용 [12]는 운전자 행동들을 준비하는 데이터셋 [28]은 7.25 시간동안의 고속도로 데이터셋 [34], [13]은 미국의 세도시를 다수의 차량들로 각 도시로의 횡단을 모은 일반인들의 데이터셋이다.

오직 회전 직진 끼어들기일 때 혼한 운전자의 행동들에 대한 추가적인 주석들을 제공한다. 게다가 CAN signals 는 각기 다른 상황아래 보행자간의 상호 작용을 한다. 최근에 [34]는 FCN-LSTM network 적용하였다. 그들은 CAN signals에 기반한 4가지 개별적 행동들을 고려했다. 운전모델 학습에 있어서. 대신에 figure6a에서 보여지는 운전자 행동의 정확한 정의를 제공한다. 여러 센서들에 의한 분류와 같은 것을 시도함으로써 멀티테스크 학습 프레임워크 은 디자인 되었다. 비슷한 아이디어는 semantic segmentation [34]에 의해 제안되었다.

#### Human Activity Understanding Datasets.

사람 행동에 대한 이해는 중요하다. 다른 데이터셋들 [10,31,15,25] 은 일찍이 제약들을 제시하도록 제안되었다. 우리의 데이터는 introduction에서 언급된 운전자 행동 학습에 가능하다. goal oriented action은 egocentric activity recognition 문제이다. [25] 데이터셋은 첫째로는 사람에 대한 egocentric activity는 운전자의 goal oriented로 정의 되고 두 번째로는 egocentric videos and signals을 학습하는 데 전달해주는 두가지 측면에서 관련이 있다.

[18,26]는 보행자에 대한 연구 데이터셋으로 [18]은 건너는 의도에 대한 [26]은 보행자들의 행동에 대한 연구이다. Robust pedestrian behavior modeling 은 또한 [16, 17]이 필요하다.

#### Visual Reasoning Datasets

Visual question answering(VQA) 인공지능의 도전적인 분야였다.A VQA agent 는 시각의 정보로 이유와 답을 낼 수 있어야한다. 많은 수의 [1, 14] 데이터셋과 [2,11] 에 대한 알고리즘들은 최근에 발표되었다. 우리의 업적으로, 결정을 만들거나 이유들을 행하거나 장면을 이해할수 있는 시스템들로 발전시킬 수 있길 바란다.

### 3.Honda Research Institute Driving Dataset

#### 3.1 Data Collection Platform

모든 센서들은 우분트 리눅스와 컴퓨터 스펙...etc.로 되어있다. 이 데이터는 2017년 1월 to 10월까지의 기록. 150GB, 104시간 1280X720 AT 30FPS로 되어있다.

#### 3.2 Annotation Methodology

운전자의 행동의 클래스를 구분짓는 것은 인식 절차와 운전자 행동의 정확한 부분 때문에 어려움이 있다. 우리의 annotation methodology는 인식 과학과 인간 요인들로 이끌어졌다. Michon [24] 운전 절차의 3가지 분류들로 제안되었다. operational processes는 차량의 조작 tactical processes는 차량과 보행자 그리고 환경에 대한 상호작용, 그리고 strategic processes 는 높은 수준의 추론, 계획 그리고 결심.

이러한 정의를 가지고, 우리는 4가지로 분류했다. goal-oriented action, stimulus-driven action, Cause and attention

goal-oriented – 회전 합류

stimulus-driven – 정지, 합류에서 나가는

Cause and attention – cause 위의 행동들에 대한 이유를 설명하는 것, attention 보행자를 도식화한 것

ELAN 라는 오픈 소스를 이용하여 비디오에 LABEL를 생성. 주석에서의 동일성을 위해서 사람의 판단으로 하나로 병합시켰고, 마지막으로 내부의 전문인에게 검토받았다.

동일성을 분석하기 위해서는 10개에 대한 외부에서 내부 전문가들의 주석과 비교한다. 이와 같은 절차로 98% 동의성을 얻었다. 그러나, 시작하고 끝시간이 평가하기에는 부족하다. 주관성을 내포하기 때문이다. 주석에 대한 동일성은 더 많은 조사가 필요하다.

### 3.3 Dataset Statistics

104시간 동안 4레이어의 구조로 주석 처리되었다. 각 137세션들이 있고 각 세션에 평균적길이는 45분이다. 통계치는 FIGURE 6B에서. FIGURE 6A는 각 행동의 예시들의 수를 보여준다. 불안정한 라벨도 이 FIGURE에서 보여질 수 있다.

## 4. Multimodal Fusion for Driver Behavior Detection

우리의 목적은 미리 정의된 클래스로 분배함으로써 세션 운전 동안 일어나는 운전자 행동들을 감지하는 것이다. 첫 번째 단계로, 우리는 GOAL-ORIENTED와 CAUSE LAYERS 에 초점을 둔다. 그리고 가능성 분배들로 해석하고 필요한 기록들을 인코딩하는 운전 상태에 학습하는 알고리즘을 만든다. LSTM 네트워크들은 이에 효율적이기 때문에 우리 모델에 중점이 되었다.

추가적으로 우리는 차량의 다이내믹에 대한 보완적인 정보들을 제공하는 보조의 신호에 접근하였다. CAN bus sensors을 이용한 figure 5 측정요소를 포함한다. 기초적인 구조는 figure 6c.에서 보여진다.

스티리밍 방식에 실제 시간에 사용되어질 수 있는 디자인 모델에 대한 필요성을 결정한다. 따라서, 우리는 우리 모델을 이전과 현재의 관찰에 기반된 만들어진 예측에 제한한다.

## 5. Experiments

이를 평가하기 위해서는, 기초 라인에 의한 실험을 시행한다. 무엇보다 우리는 지역 기반으로 데이터를 쪼갬다. 이 방법으로 같은 지역에 훈련되어있는 곳을 테스트하는 것을 피한다. figure 3은 training (in blue) and test(in red) splits.

우리는 베이스라인 모델을 위해 can bus sensors을 이용한 값과 비디오로부터 프레임을 얻는다. Imagenet[5]으로 얻어진 Conv2d\_7b\_1x1 layer of inceptionresnet-v2 [32] 이용했다.

정제하지 않은 센서 값들은 visual features의 연속 전에 fully-connected layer을 거친다. 결국엔, 격자인 공간으로 cnn에 의해 visual features은 나타내어졌다. table2 에 cnn 으로 인해 공간적인 문제는 해결된다. LSTM 모든 실험에서 2000번 실험으로 설정되었다. 이전 LSTM 은 다음 단계의 LSTM으로 사용되었다. 학습은 시간에 잘라진 BACKPROPAGATION을 통해 이루어진다.

GOAL, CAUESE 레이어를 위해, 우리는 개별적 LSTM들을 학습시켰다. BATCH SIZE 클수록 수렴성이 향상됨을 확인했다. 우리는 LSTM을 0.9로 dropout keep probability를 설정한다. 앞전 이전배경 사이에 불안정한 데이터와 스스로의 운전 분류를 고려하면서 우리는 [20] 엔트로피 조정을 이용하였다.

우리의 평가 전략은 행동 감지 부분으로부터 영감이 되어졌다.[30] 라벨로 되어지는. 특히, [30]은 confidence 점수에 따라 모든 프레임에 랭킹을 매기면서 labeling과 average precision 산출했다. 테스트는 지형 분할을 거쳐서 27만4천 프레임들에서 3fps 당 ap 평가의 평균의 합을 포함한 37세션을 포함한다.

## 6. Results and Discussion

요약하다가 이 부분이 주요 기술적 제시 부분이라 꼼꼼히 읽으시는게 나을듯해서 해석 안해부렸어요~ :)

## 7. Conclusion

우리의 실험 목적은 자극하기위해 운전자 행동을 인지하는 알고리즘을 제안하는 회사들. 이를 가능하기위해 우리는 운전자 행동을 분해하는 Goal-oriented, stimulus-driven, cause and attention. 4가지를 제안한다. 최초의 결과들은 이 업무가 rgb 프레임에 기반된 인식되는 방법은 어렵다는 것을 보여준다. 센서들이 추가되면 정확성이 늘어나지만, 더 나은 시간적 모델링, 학습 전략이 필요하다. 행동들과 다른레이어 들간의 관계를 파악하기 전에, ex. 운전자와 상황들에 대한 관계.