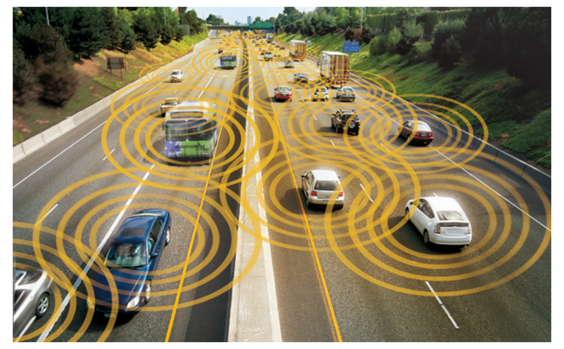
DBUS: Human Driving Behavior Understanding System

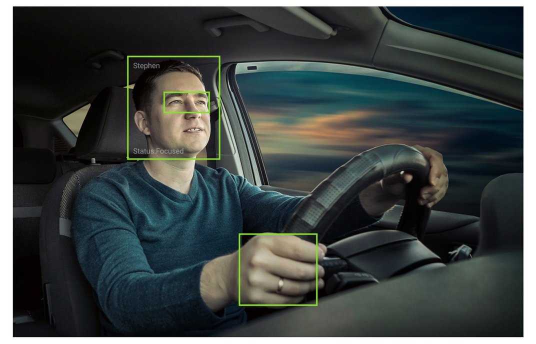
Max Guangyu Li[‡*], Bo Jiang[†*], Zhengping Che[†], Xuefeng Shi[†], Mengyao Liu[†], Yiping Meng[†], Jieping Ye[†], Yan Liu[‡]

[†]Didi Chuxing [‡]University of Southern California

Introduction

- What is human driving behavior understanding?
  - Understand ***how humans drive and interact with environments***
- Driving behavior understanding in intelligent transportation systems
  - Autonomous driving (AD)
  - Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I)
  - Driving safety monitoring system – *Top priority for ride-sharing platform & fleet management*



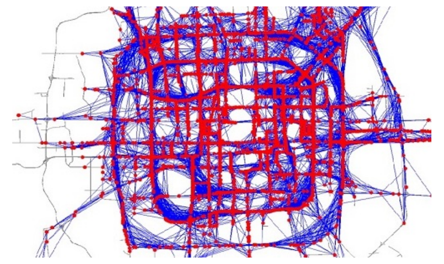


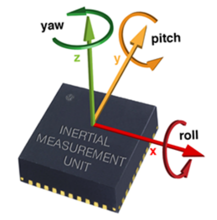
- Why human driving behavior understanding is **essential**?
  - AD: Proposing human-like P&C strategies, etc.
  - V2V/V2I: Providing insight for an efficient and convenient design of systems, etc.
  - Driving safety monitoring system: Real-time driving safety monitoring, etc.
- Key ***challenges*** come from
  - Sophisticated real-world traffic scenarios – *Environment*
  - Huge diversities of different driving styles – *Drivers*

Motivation

- Collection of human driving data on ride-sharing platform
  - ***Heterogeneous*** – including videos, GPS/IMU signals, etc.
  - ***Large-scale*** – offering potential for in-depth analysis

- Major challenges of human driving understanding
  - Handling different complex tasks
  - Multi-level analyzation
  - Heterogeneous data source
  - System efficiency
- Existing works for human driving behavior understanding
  - Only focus on relatively small dataset
  - Only analysis it at single level or on subset of tasks
  - Only leverage single type of data





- Our proposed solution
  - DBUS: Human Driving Behavior Understanding System

Problem Formulation

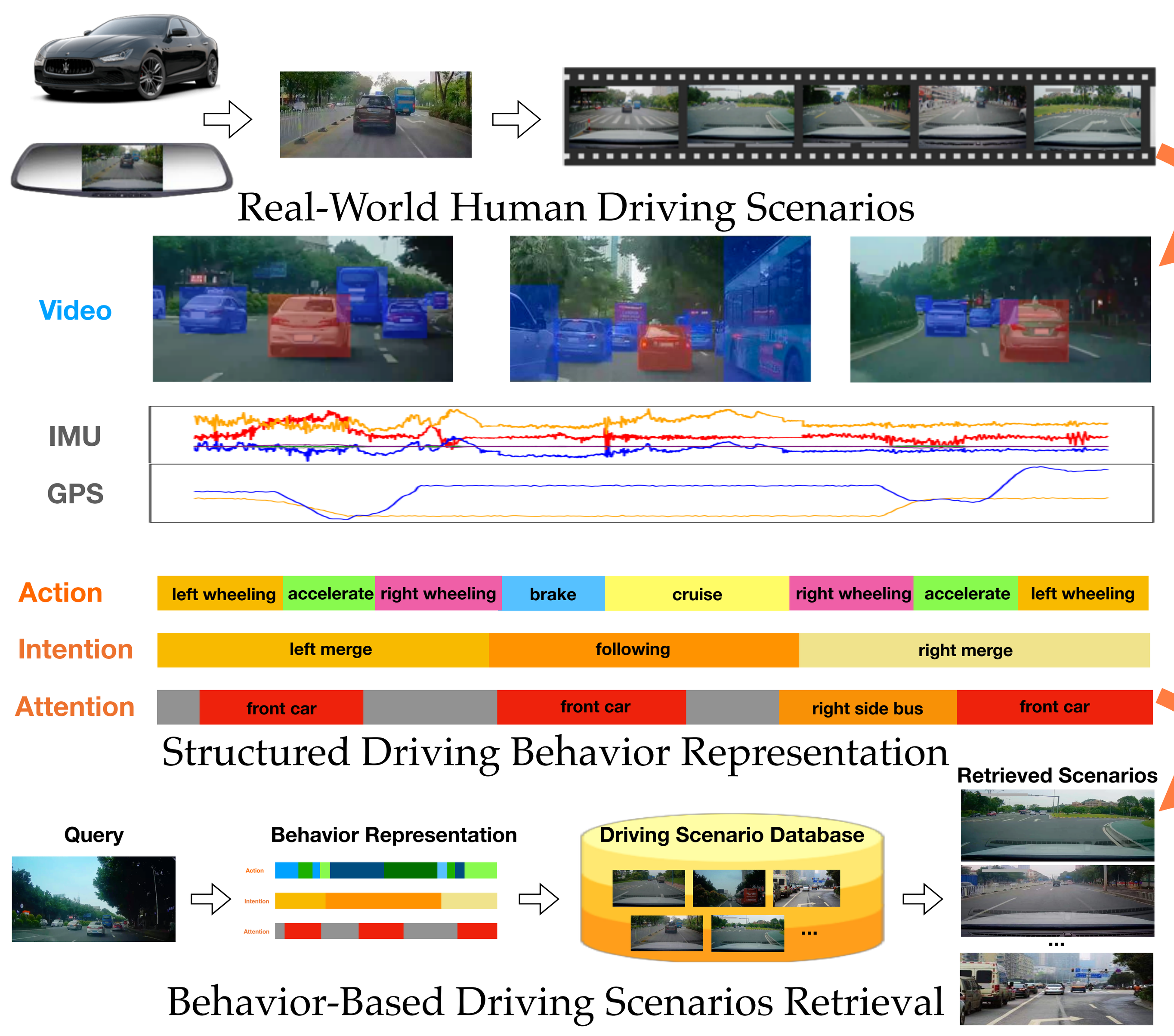
- **Input** – Human driving scenario $\mathcal{D} = (\mathbf{V}, \mathbf{S})$ with time horizon T
  - $\mathbf{V} = \{\mathbf{v}_t\}_{t=1}^{T}$ refers frames from front-view camera
  - $\mathbf{S} = \{\mathbf{s}_t\}_{t=1}^{T}$ denotes GPS/IMU signals
- **Output** – Three-level structured representation of driving behavior $\mathcal{B} = (\mathbf{M}, \mathbf{W}, \mathbf{A})$
  - $\mathbf{M} = \{\mathbf{m}_t\}_{t=1}^{T}$ refers pre-defined basic driving actions
  - $\mathbf{W} = \{\mathbf{w}_t\}_{t=1}^{T}$ refers pre-defined driver's intentions
  - $\mathbf{A} = \{\mathbf{a}_{mask}^t, \mathbf{a}_{obj}^t\}_{t=1}^{T}$ denotes driver's attention object categories

References

1. AB Ellison, et al. *Driver characteristics and speeding behaviour*. In Australasian Transport Research Forum (ATRF), 33rd, 2010, Canberra, ACT, Australia, 2010.
2. Dariu M Gavrila. *Pedestrian detection from a moving vehicle*. In European conference on computer vision, pages 37–49. Springer, 2000.
3. Jiangpeng Dai, et al. *Mobile phone based drunk driving detection*. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), pages 1–8. IEEE, 2010.
4. Minh Van Ly, et al. *Driver classification and driving style recognition using inertial sensors*. In Intelligent Vehicles Symposium (IV), 2013 IEEE, pages 1040–1045. IEEE, 2013.
5. Luis M Bergasa, et al. *Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors*. In Intelligent Vehicles Symposium Proceedings, IEEE, 2014.
6. Pengyang Wang, et al. *You are how you drive: Peer and temporal-aware representation learning for driving behavior analysis*. In Proceedings of the 24th ACM SIGKDD, pages 2457–2466. ACM, 2018.

Overview

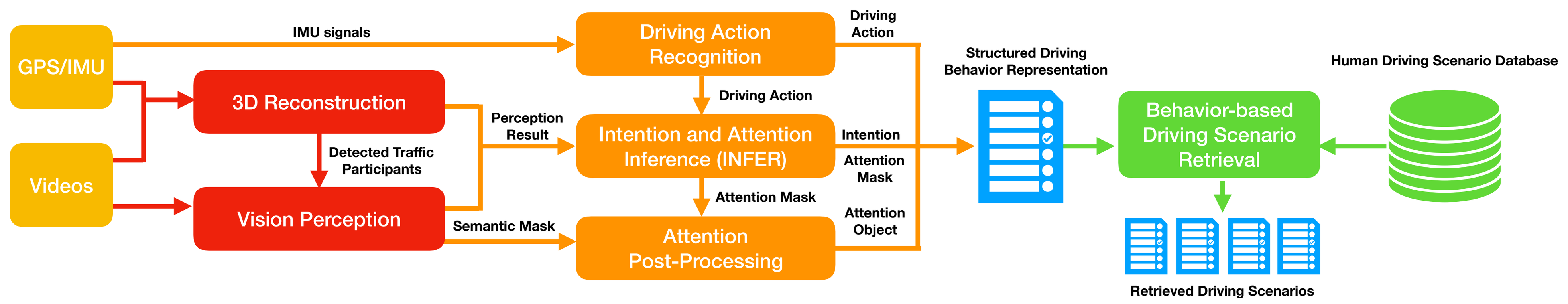
- System overview of *DBUS*



- Major tasks & applications
  - Driving behavior analysis – Joint inference of $\mathbf{W}$ and $\mathbf{A}$
  - Driving scenario search & retrieval – Given $\mathcal{B}$, retrieve top-K relevant $\mathcal{D}$ from a *massive* database

System Architecture

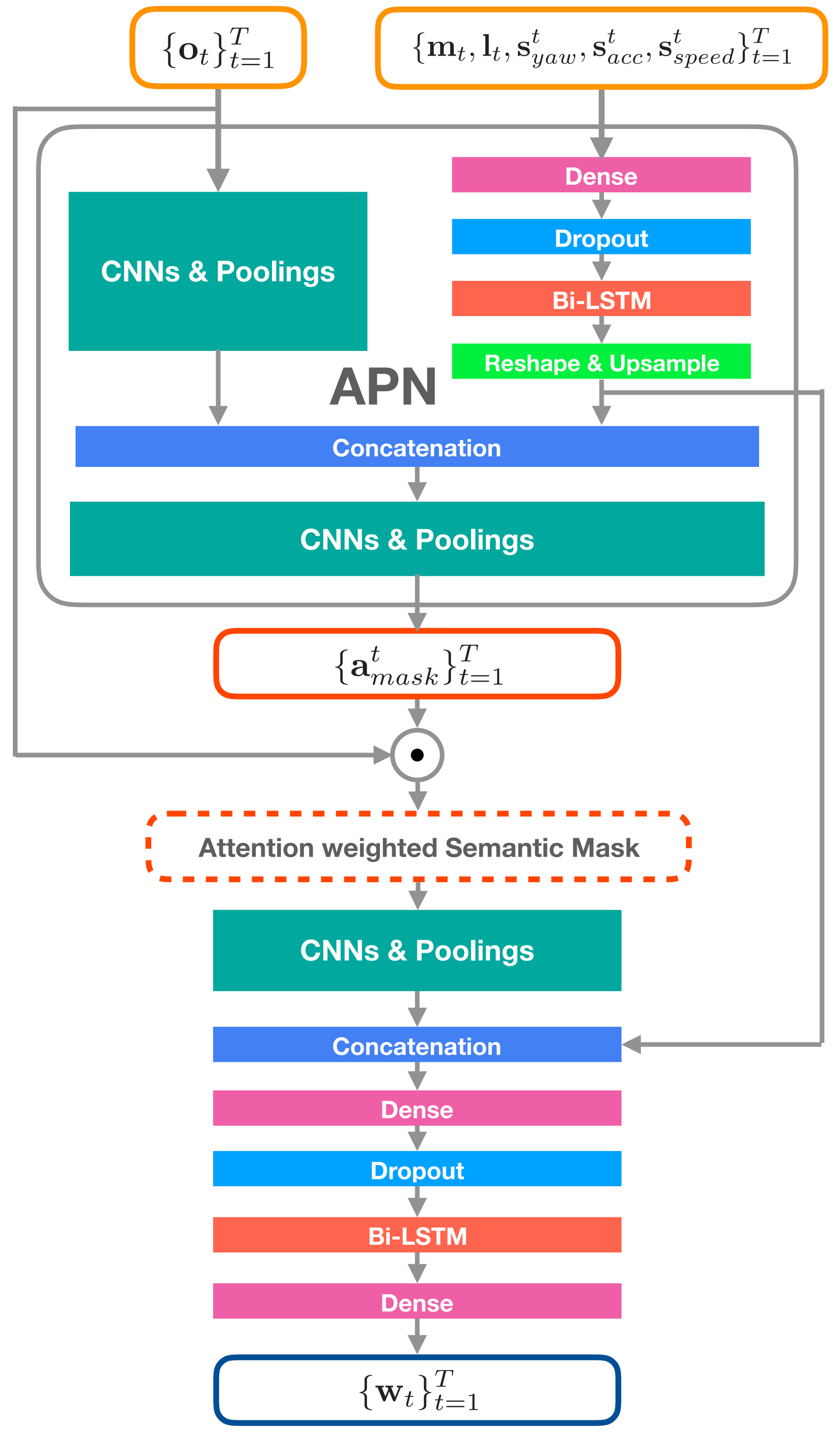
- System architecture of *DBUS*



- Major modules of *DBUS*
  - **Perception** – Vision perception and 3D reconstruction of driving scenarios, etc.
  - **Driving Behavior Analysis** (core module!!!) – Generates 3-level *structured representation*
  - **Driving Scenario Retrieval** – Efficient behavior-based retrieval of relevant driving scenarios
- Workflow of *DBUS*
  1. Feed $\mathcal{D}$ -> **Perception** -> $\mathcal{P} = (\mathbf{O}, \mathbf{D}, \mathbf{L})$
     - $\mathbf{O}$: Semantic masks of detections
     - $\mathbf{D}$: Distance between ego-vehicle and the nearest front vehicle
     - $\mathbf{L}$: Ego-vehicle's relative location on the road based on the lane perception
  2. Feed $\mathcal{P}$ + GPS/IMU -> **Driving Behavior Analysis** -> $\mathcal{B}$
  3. **Driving Scenario Retrieval** takes $\mathcal{B}$ and returns top-K relevant $\mathcal{D}$

Driving Behavior Analysis

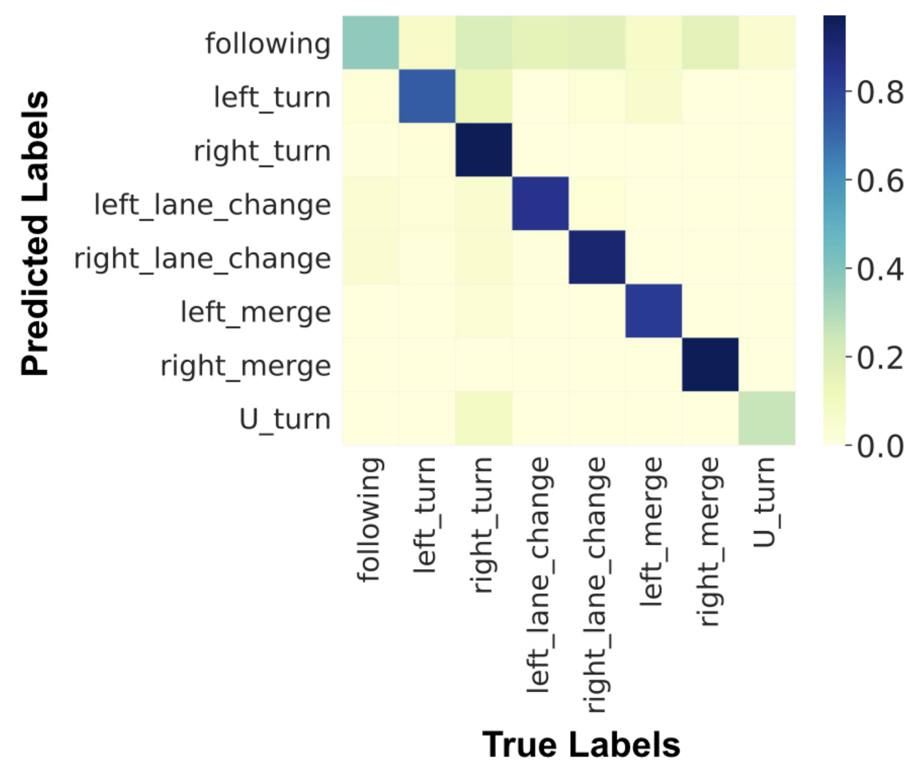
- Basic driving action inference
  - Based on $\mathbf{S}$ with a rule-based manner
- Intention & attention inference
  - Introduced a deep model INFER
    - ∗ Attention proposal network (APN)
    - ∗ Intention inference network
  - **Input** – Set of features
    - ∗ $\{\mathbf{o}_t, \mathbf{d}_t, \mathbf{m}_t, \mathbf{l}_t, \mathbf{s}_{yaw}^t, \mathbf{s}_{acc}^t, \mathbf{s}_{speed}^t\}_{t=1}^{T}$
  - **Output** – $\{\mathbf{w}_t\}_{t=1}^{T}$ & $\{\mathbf{a}_{mask}^t\}_{t=1}^{T}$
    - ∗ We use $\{\mathbf{a}_{mask}^t\}_{t=1}^{T} + \{\mathbf{o}_t\}_{t=1}^{T}$ to get $\{\mathbf{a}_{obj}^t\}_{t=1}^{T}$



- Attention mask prediction and intention prediction

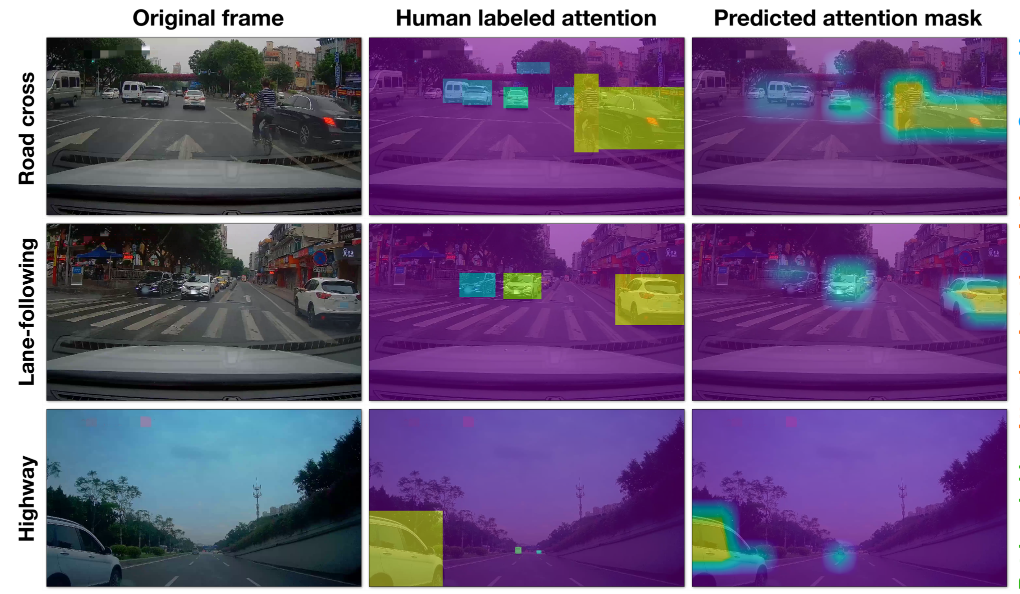
	MSE (Attention masks)	**ACC** (Intentions)
SVM	-	0.193
XGBoost	-	0.258
INFER-NO-SM	-	0.276
INFER-ONLY-SM	0.032	0.693
INFER-NO-ATTN	-	0.628
INFER	**0.025**	**0.772**

- Confusion matrix of intention prediction



Case Study

- Attention inferred by *DBUS*



- Retrieved driving scenario