Creating Affordable and Reliable Autonomous Vehicle Systems

PERCEPTION INSIGHT INTELLIGENCE

Shaoshan Liu

shaoshan.liu@perceptin.io

Autonomous Driving

Sensing
- GPS/IMU
- LiDAR
- Camera

Perception
- Localization
- Object Recognition
- Object Tracking

Decision
- Path Planning
- Action Prediction
- Obstacle Avoidance

Operating System

Hardware Platform

↔

Cloud Platform
- HD Map
- Model Training
- Simulation
- Data Storage

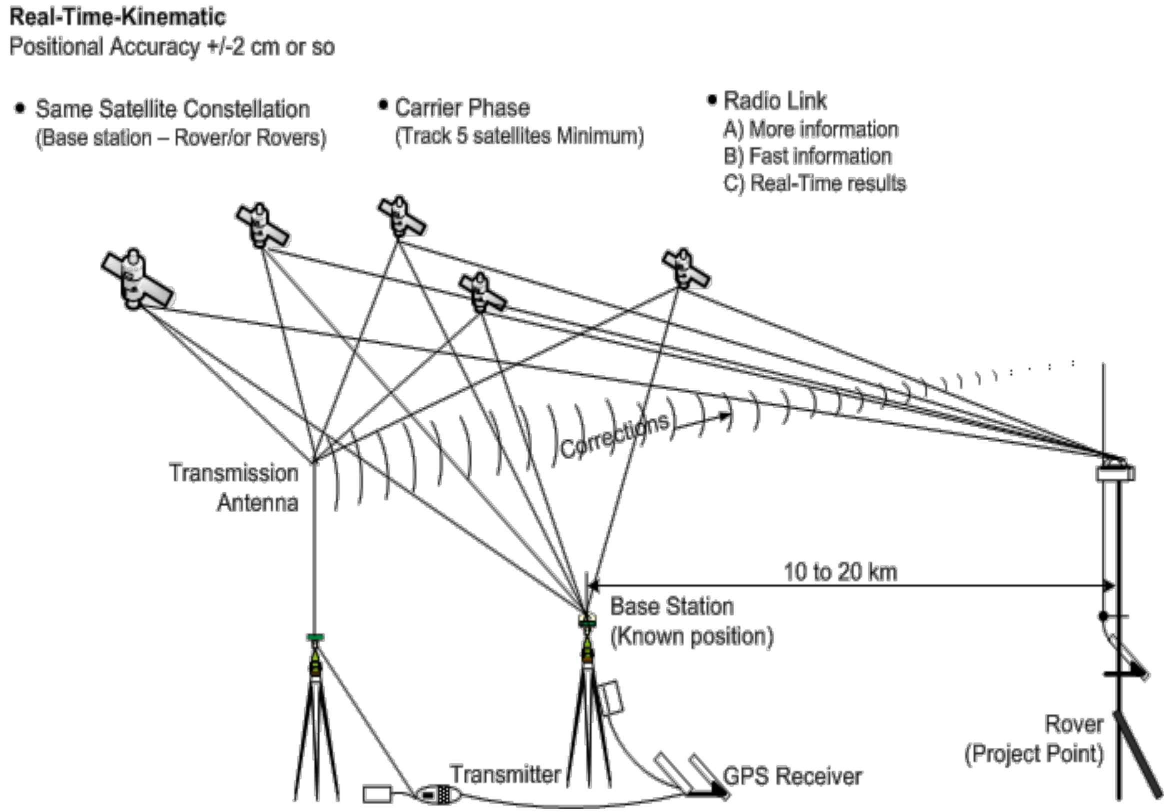
Localization

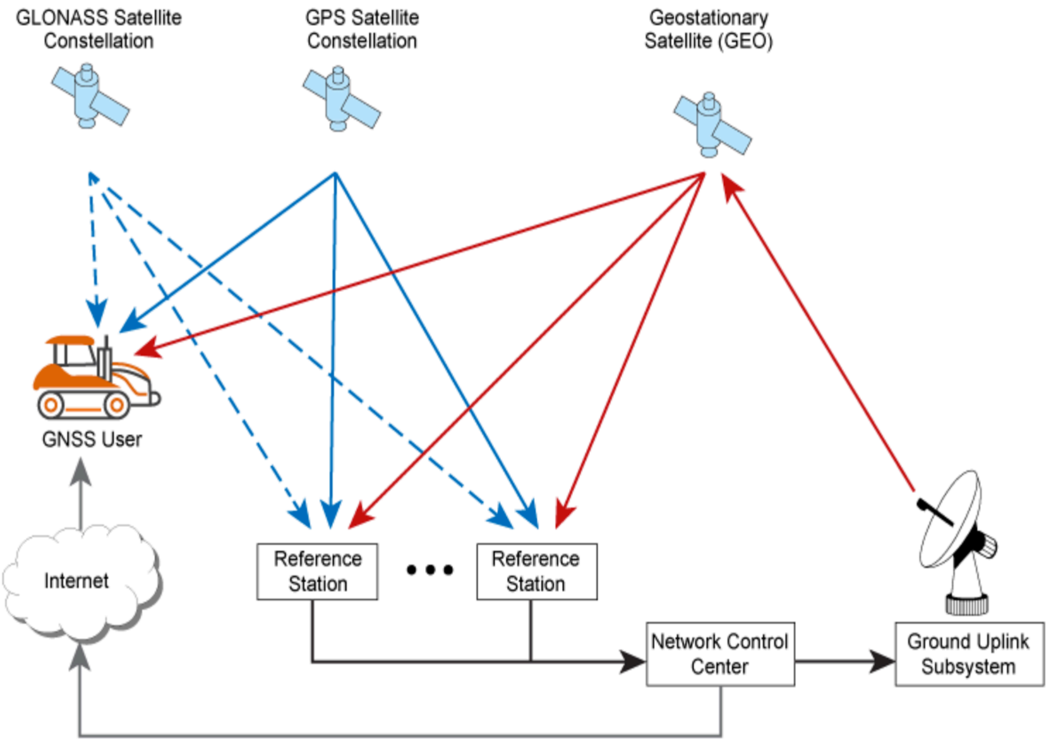
Most crucial task of autonomous driving

Solutions: GNSS but with variations, LiDAR, INS, Vision, Odometry

Localization: GNSS/INS

Contributing Source	Error Range
Satellite Clocks	±2 m
Orbit Errors	±2.5 m
Inospheric Delays	±5 m
Tropospheric Delays	±0.5 m
Receiver Noise	±0.3 m
Multipath	±1 m

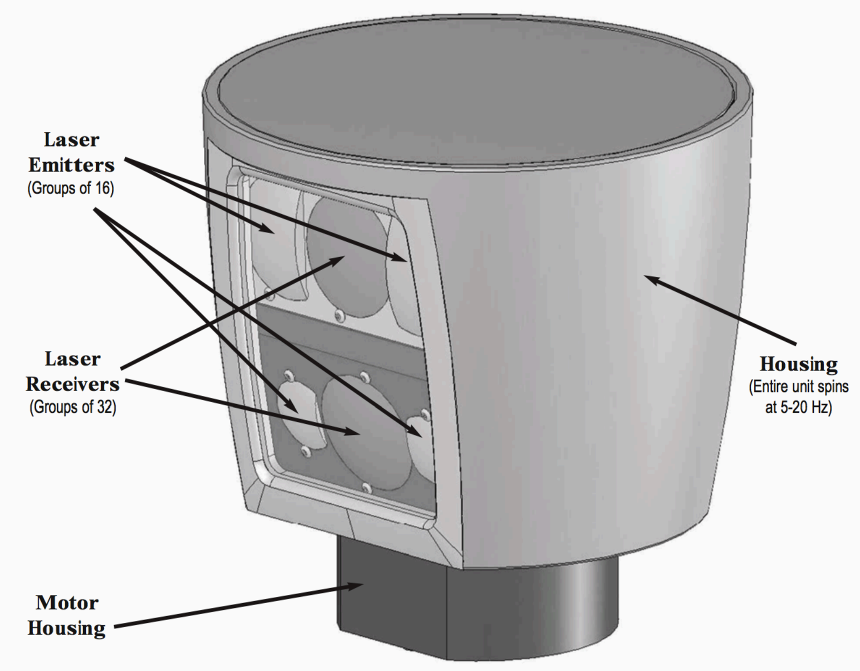


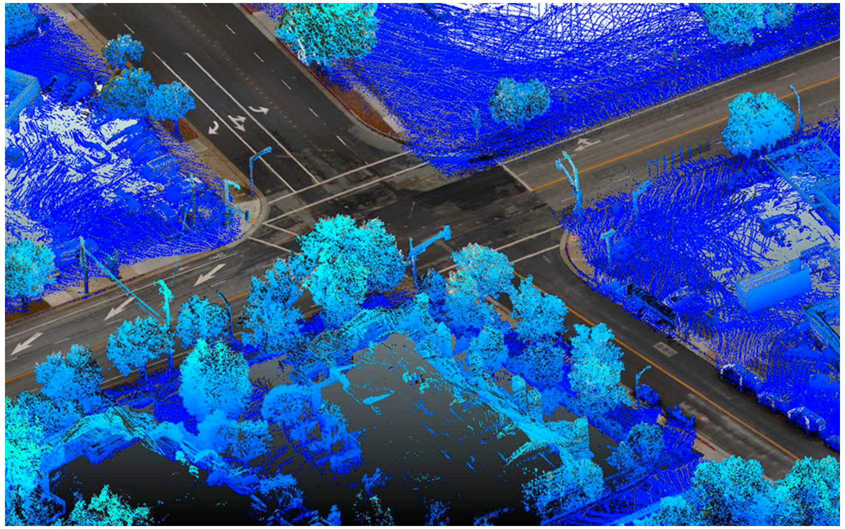


Localization: LiDAR and HD Map

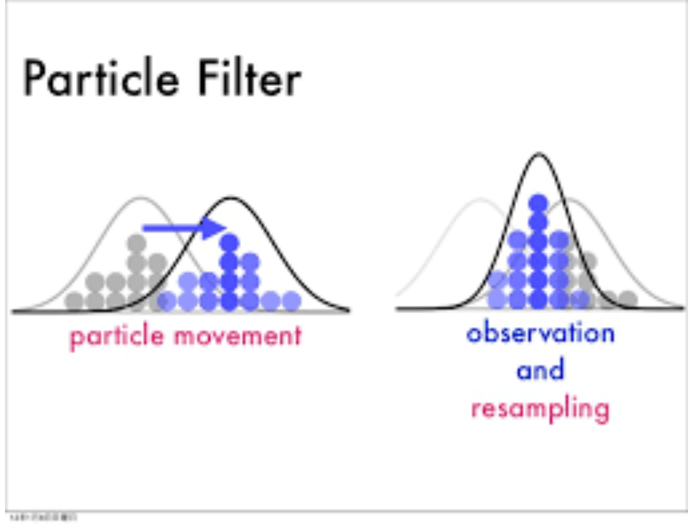
Captures a 3D environment

HD Maps





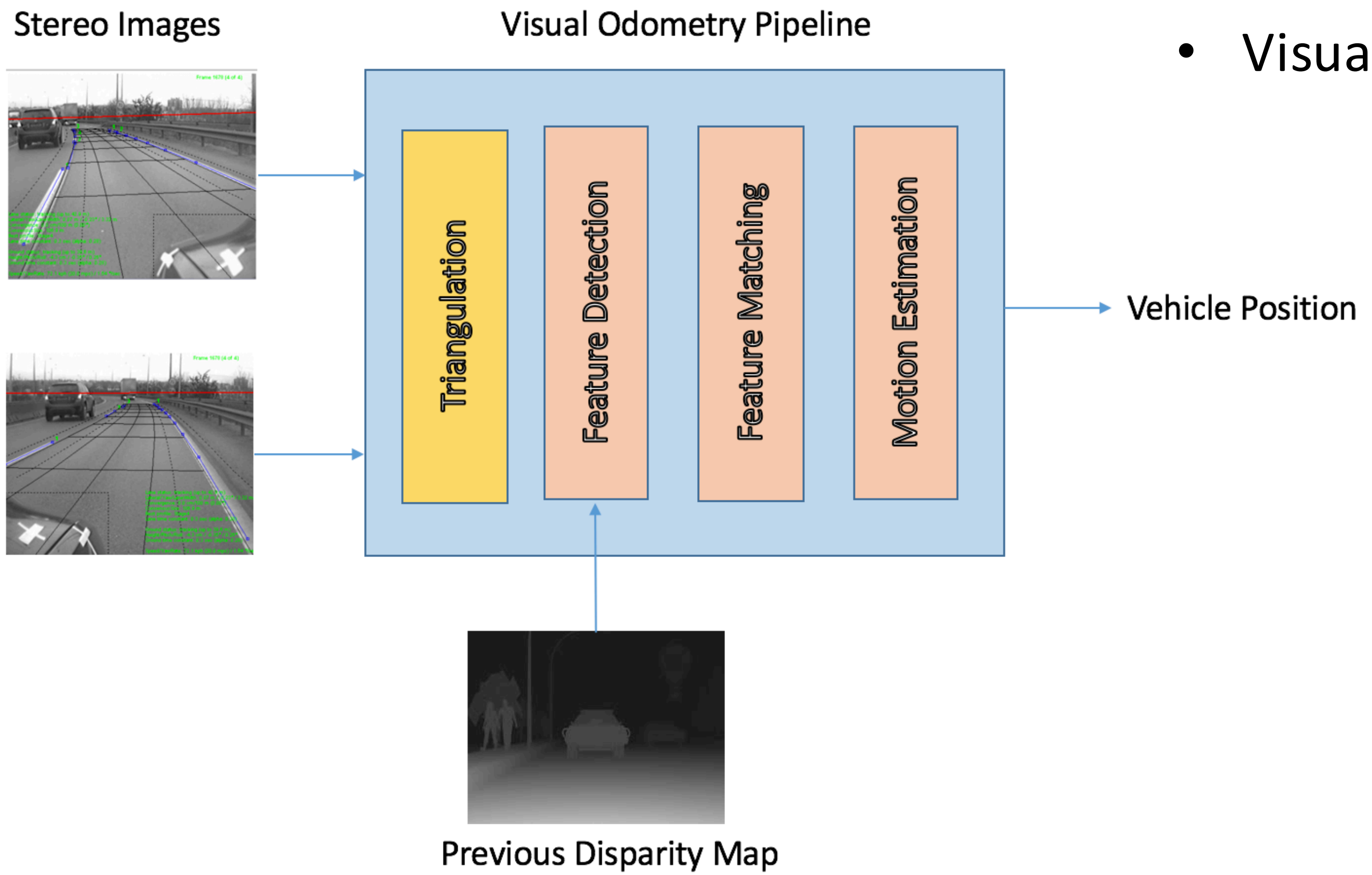
Compare LiDAR scan to HD Map



1. **Generate a set of particles randomly distributed in space**
2. **For each particle, calculate the probability of it located at the current location of the vehicle**
3. **Pick the one with the highest score and use it as the vehicle location**

Localization: Visual Odometry

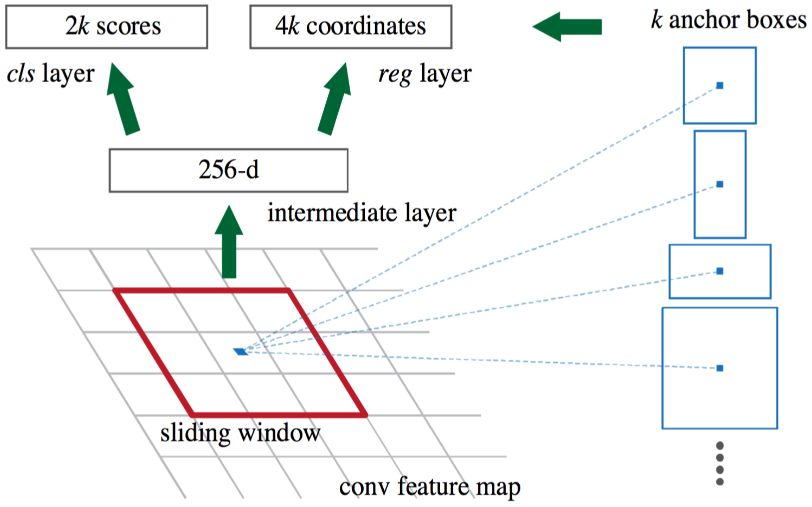
- Stereo visual odometry
- Mono visual odometry
- Visual inertial odometry

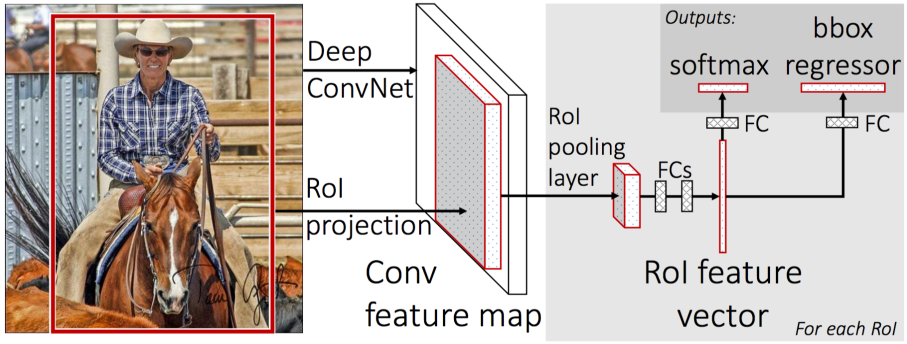


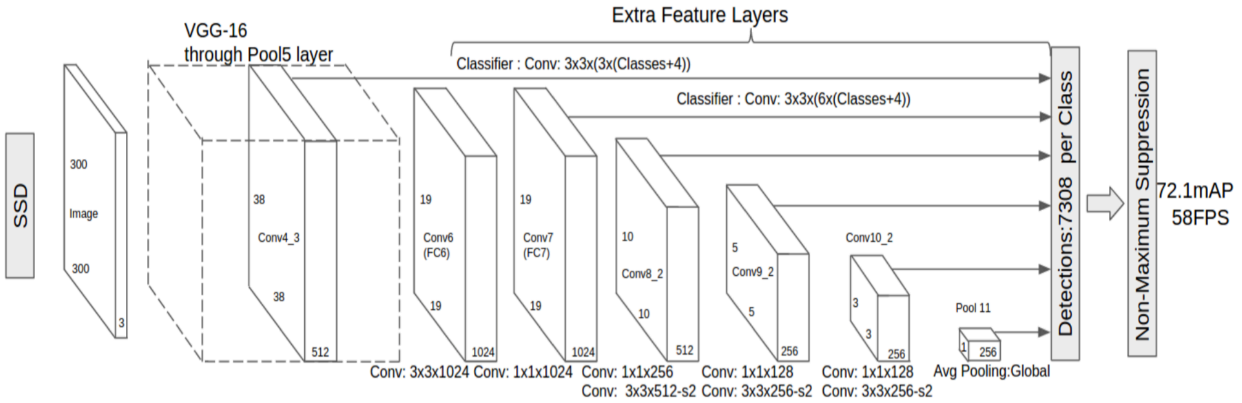
Perception

- ***Understanding of the environment***
- ***Pedestrian, Cyclist, Vehicle recognition***
- ***Road structure recognition***
- ***Traffic lights identification***
- ***Detection of moving objects, etc.***

Detection: Faster R-CNN and SSD







Faster R-CNN

1. **Obtain ROI**
2. **Perform Classification**

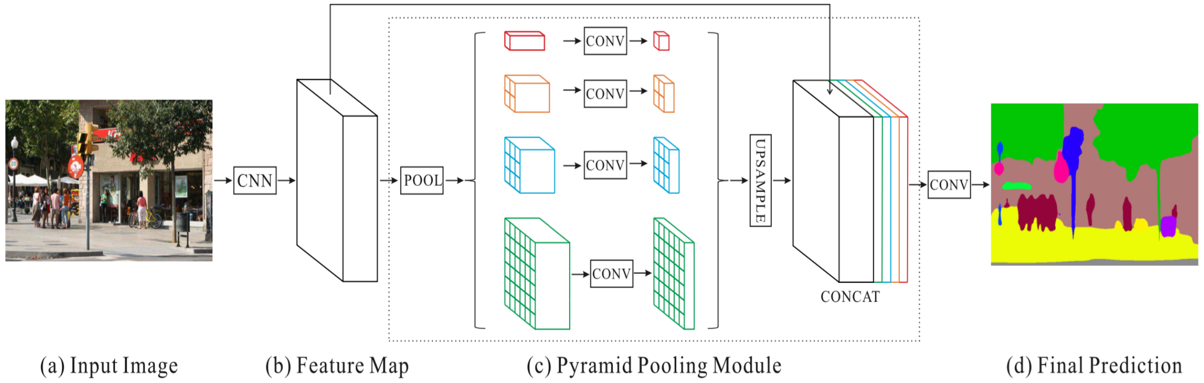
High accuracy but too slow for E/S

SSD: Single Shot MultiBox Detector

Generates object in one pass

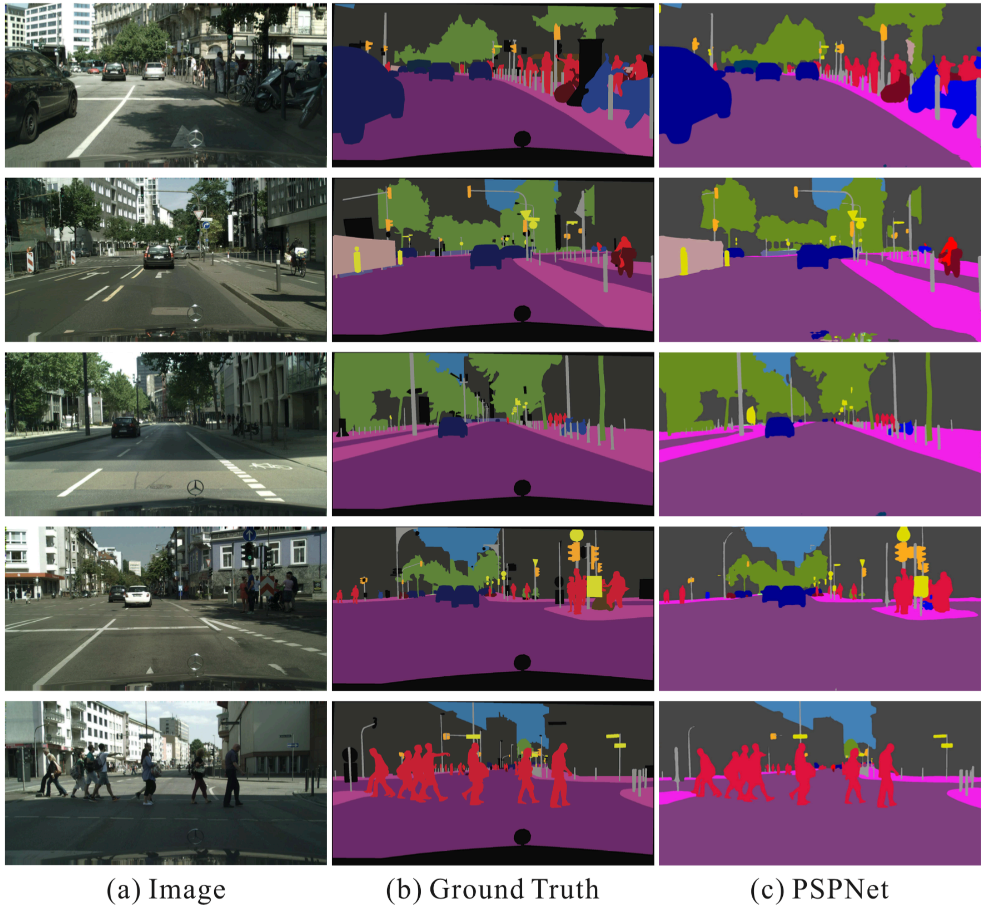
High accuracy, faster, but expensive

Semantic Segmentation

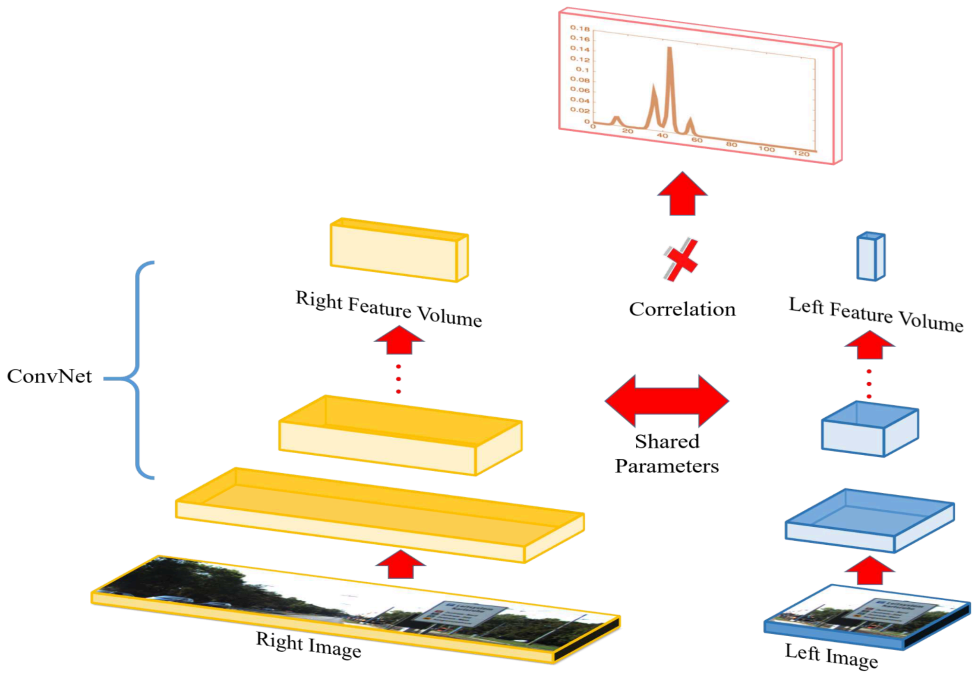


PSPNet: Pyramid Scene Parsing Network

1. **Process input images**
2. **Generate feature map**
3. **Pyramid pooling to reduce spatial resolution**
4. **Concatenate similar features into different segments**
5. **Generate final prediction**



Stereo Matching: content CNN



- **Two branches of convolution layers**
- **Share weights**
- **One for each image**
- **Outputs are joined to generate results**

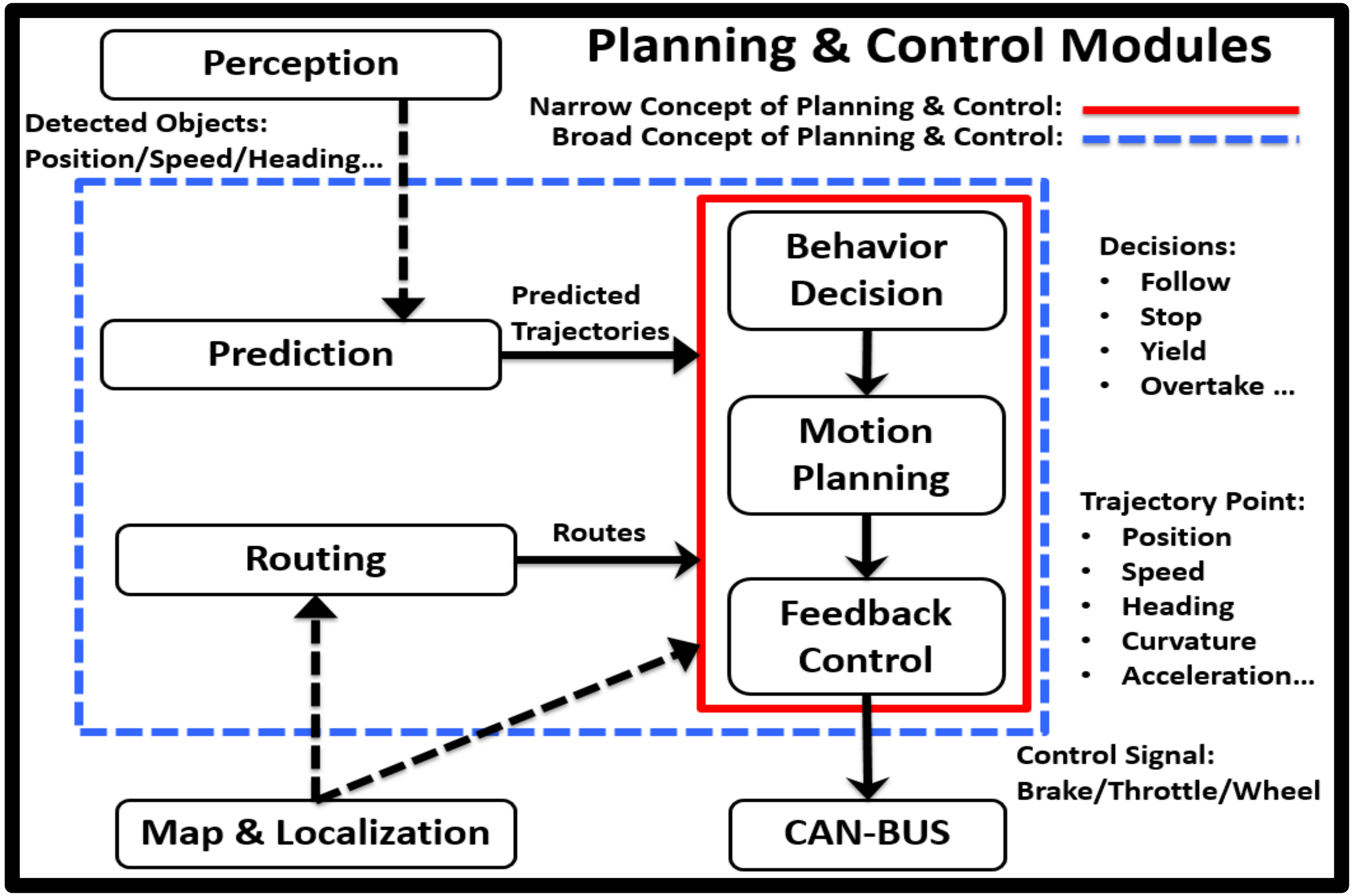
Matching error reduced by 50%
(compared to Semi-Global block Matching)

High Computation Costs

Planning and Control

Decision Making, a.k.a. “The Brain”

High-Level Architecture of the P&C Pipeline



Perception: pedestrian detection and tracking, etc

Localization and mapping: real-time position (lane-level)

CAN-bus: connects to control (Controller Area Network)

Prediction: traffic prediction

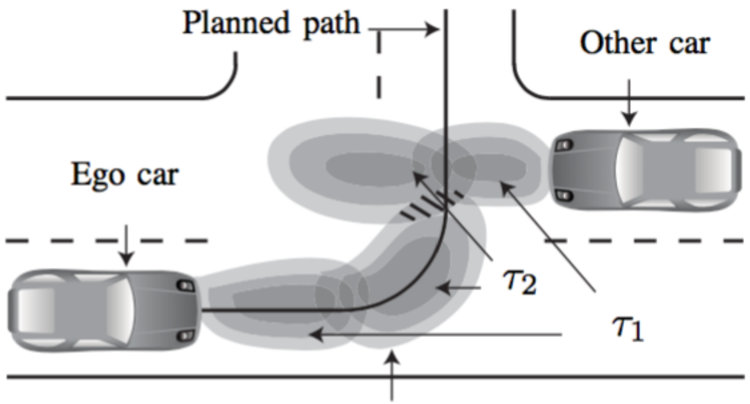
Routing: how to get from A to B

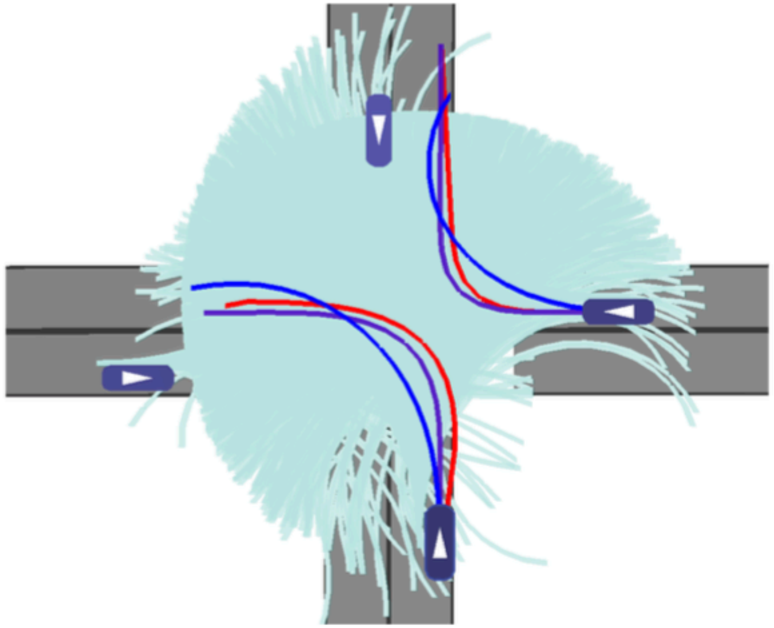
Behavior decision: high-level behavior

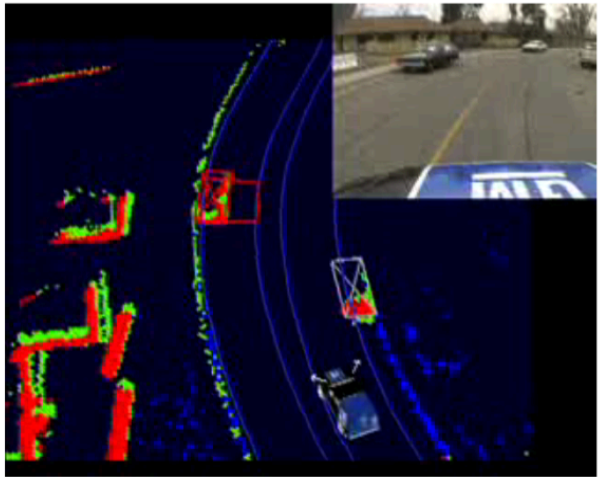
Motion planning: detailed action plans

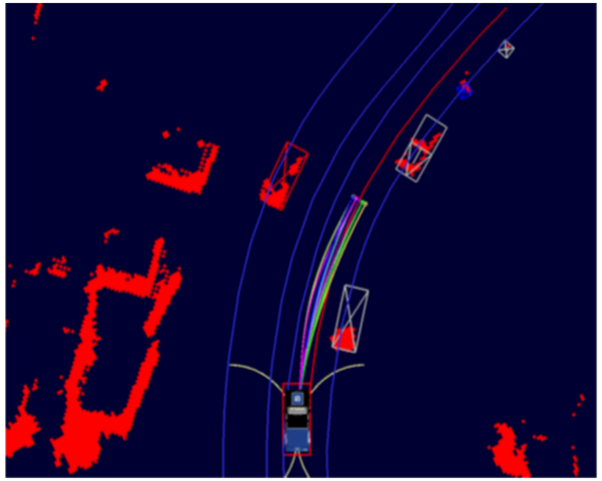
Feedback and control: generates detailed control plans

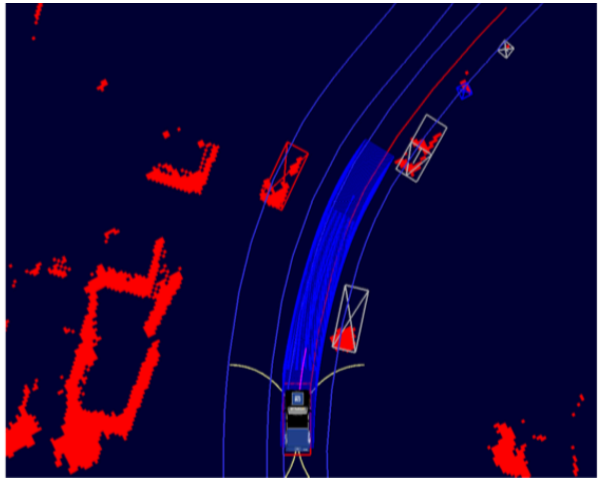
Action Prediction





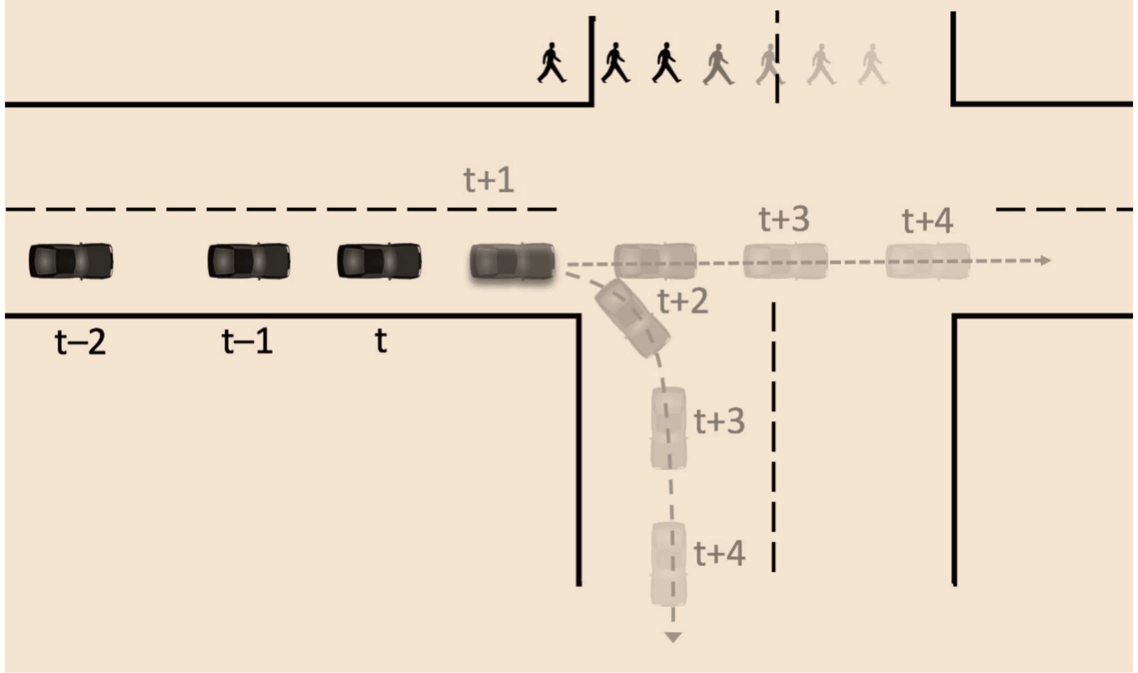


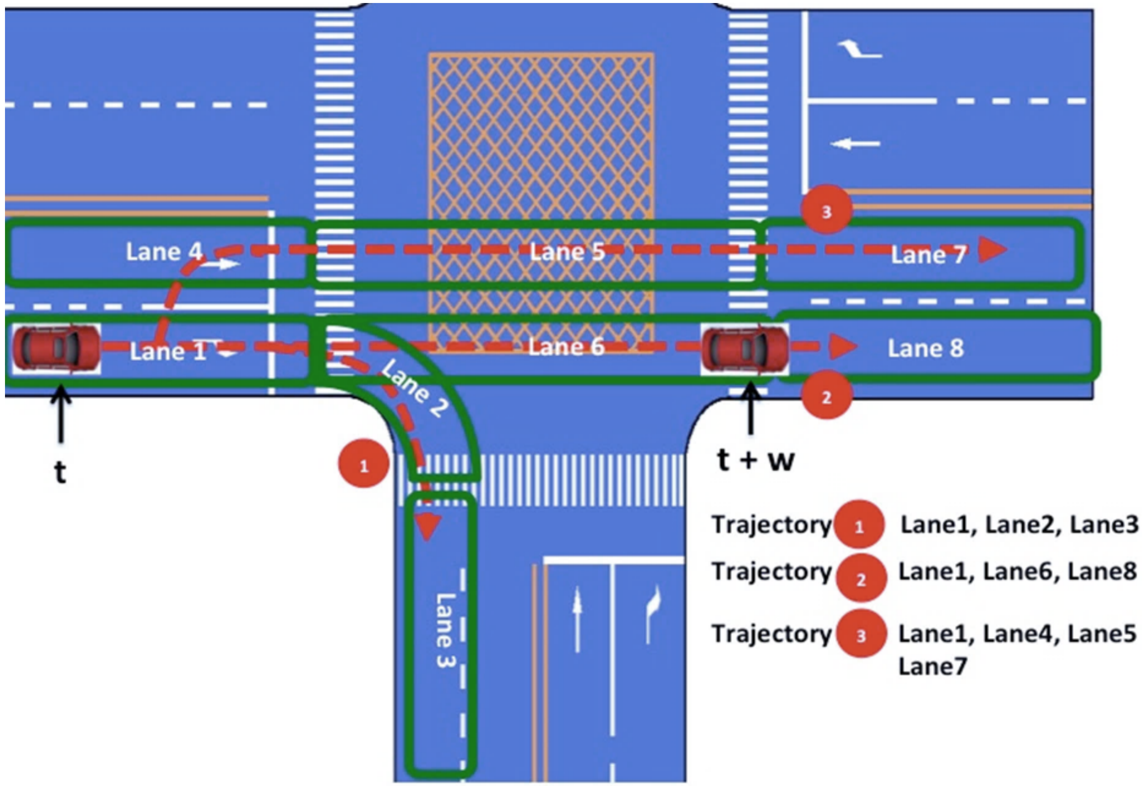




Traffic Prediction

- ***Classification problem*** **for categorical road object behaviors**
- ***Regression problem*** **for generating the predicted path with speed and time info**

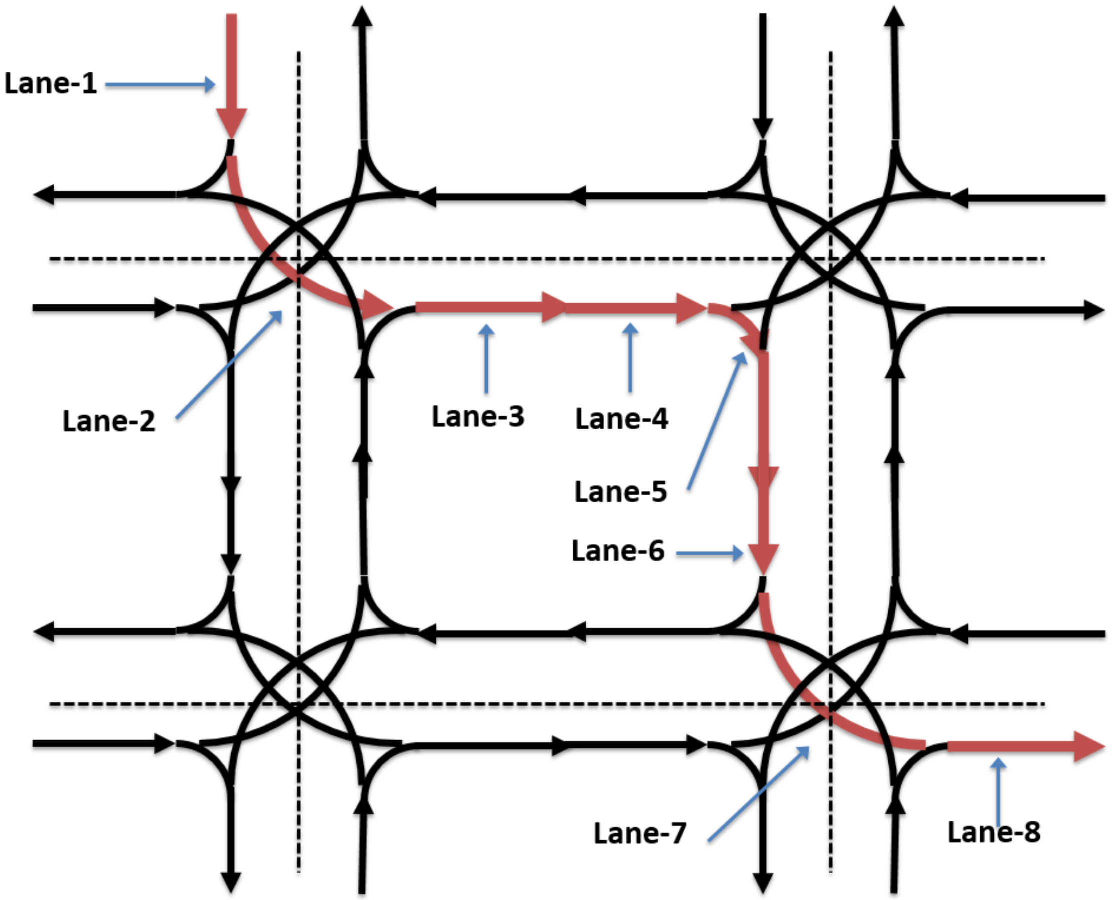


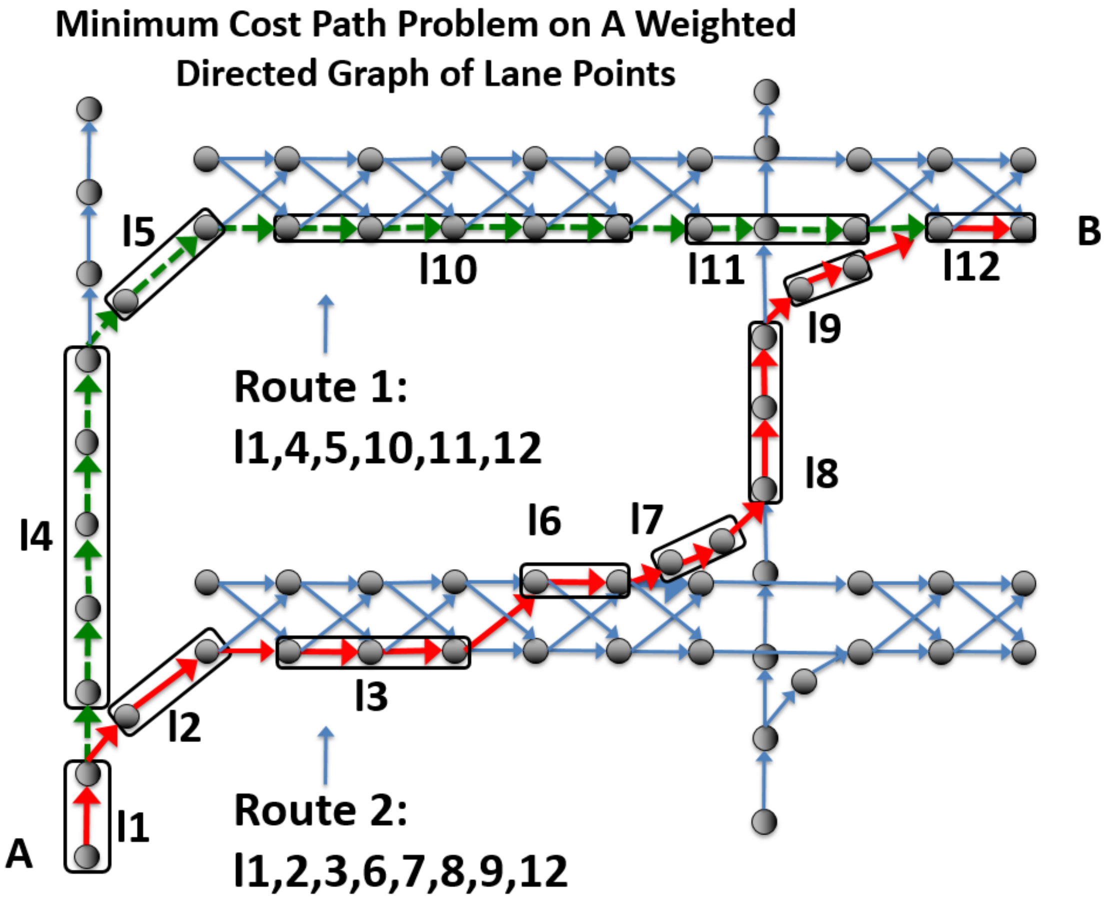


Lane-Level Routing

- **Similar to Google Maps routing**
- **Shortest path problem: *Dijkstra* and *A****

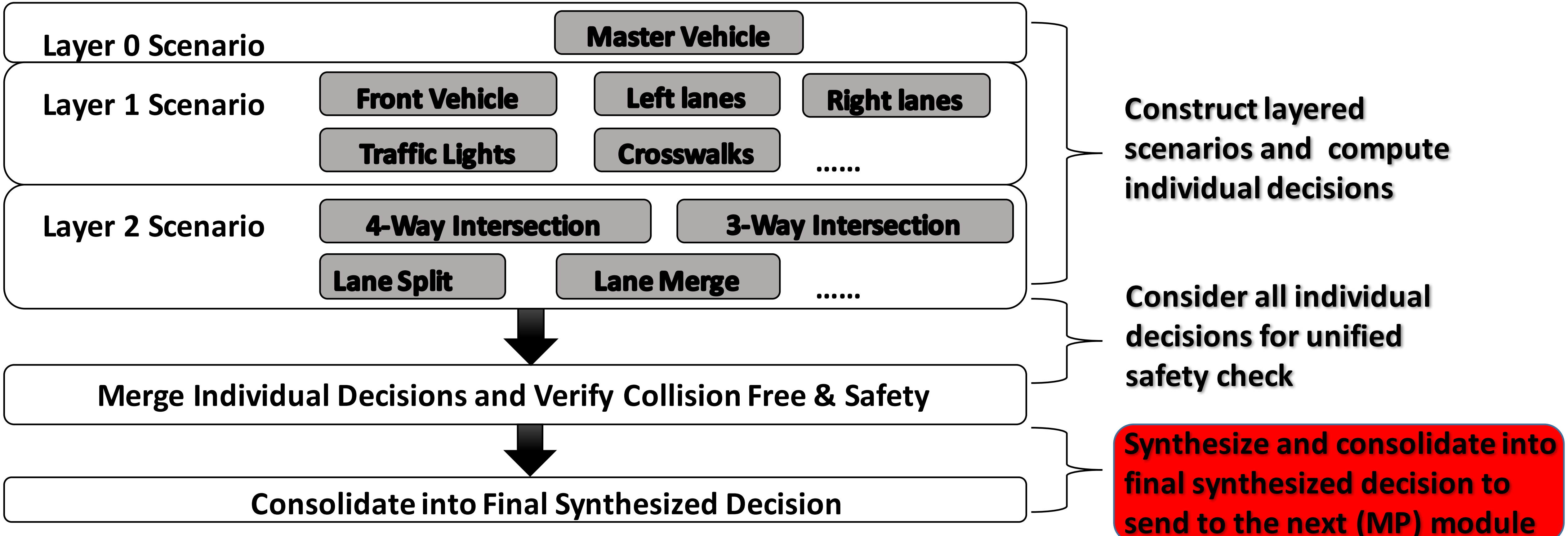
Two Algorithms for Lane-Level Routing





Behavioral Decisions - Layers

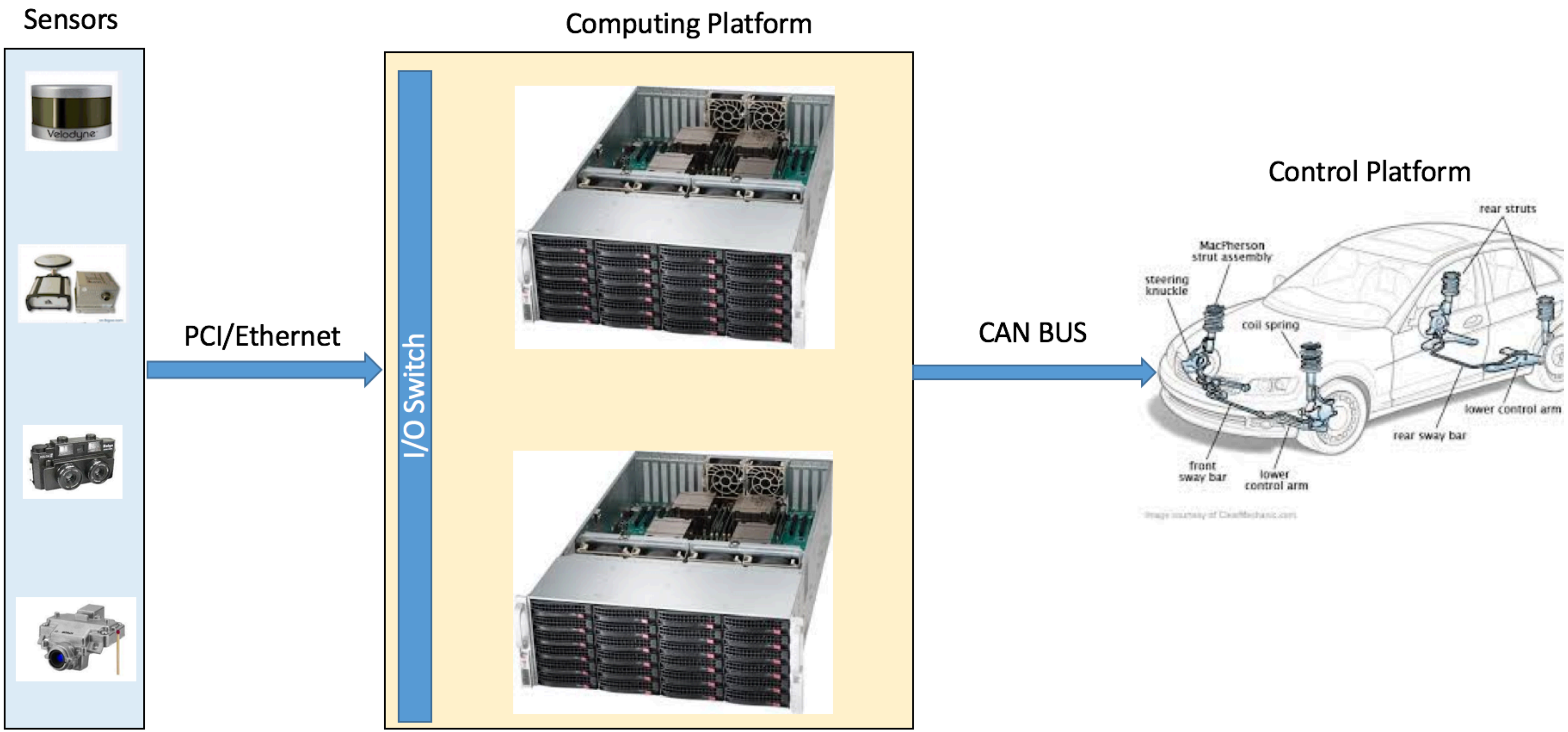
- **Ruled-based “divide-and-conquer” approach: layered scenarios**
- **Markov Decision Process**
- **Synthesized decision and individual decisions**



Client Systems

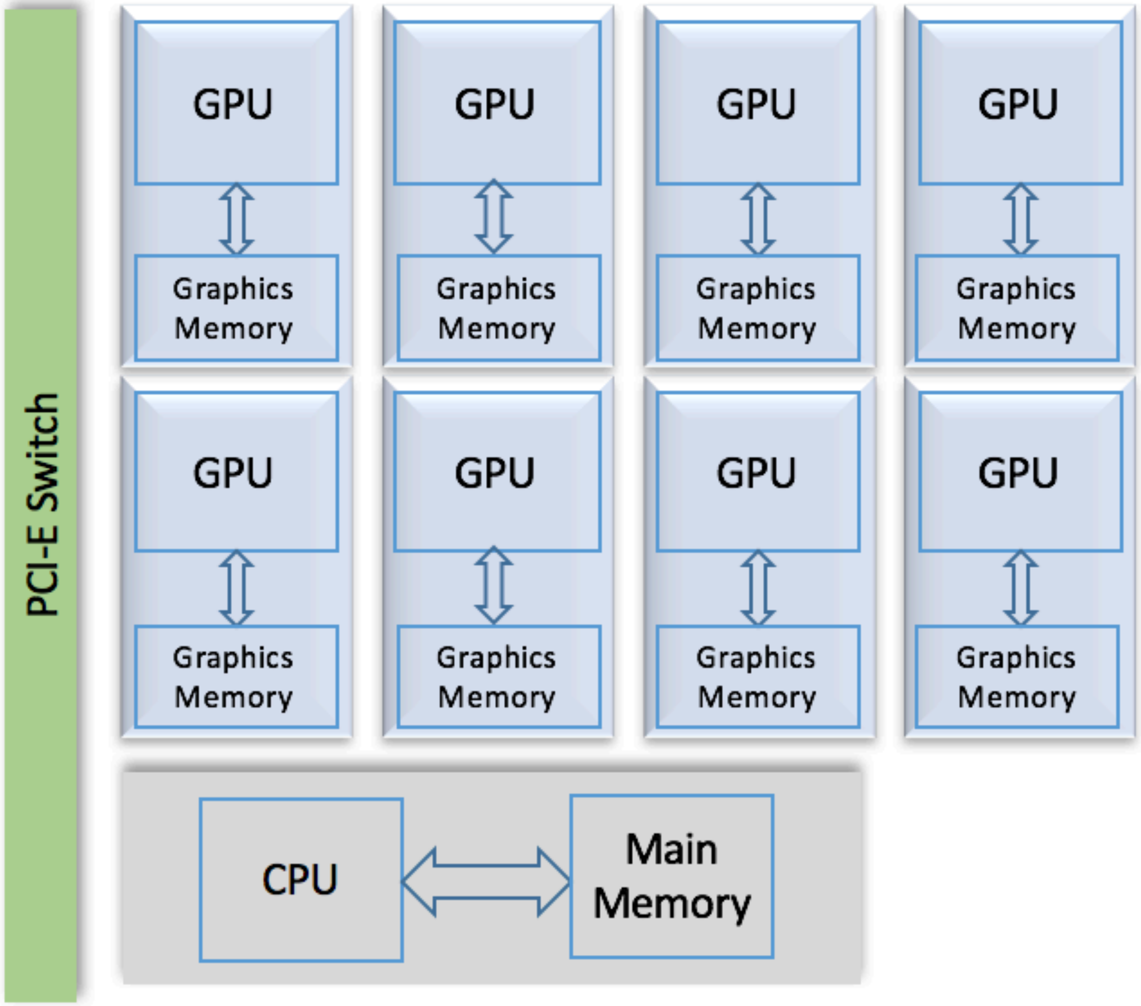
Robustly and reliably combining all these modules onto physical hardware

Hardware Platform



Hardware Platform

- High Performance
  - CPU + 8 ~ 16 GPUs
  - 60 TOPS/s
- High-Power Consumption
  - 3000 W at peak
- High Cost
  - $20000 ~ $30000
- Heat Dissipation
  - Special fan design needed

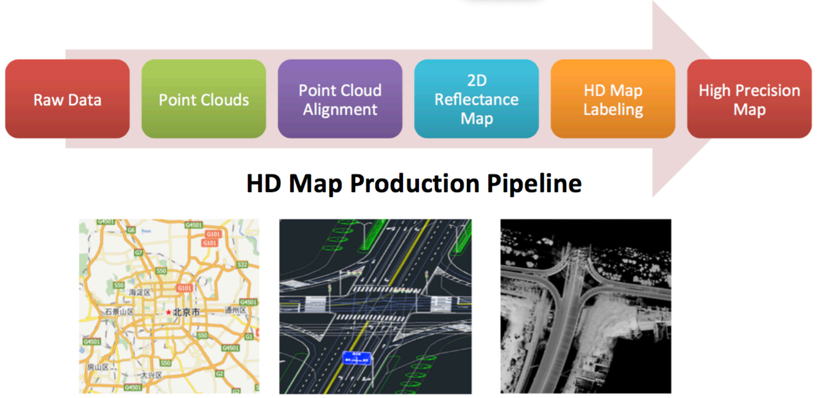


Affordability

Cost Breakdown

> $100,000 USD Sensing Hardware Cost

> $10,000 USD Computing Hardware Cost



millions of USD to create a maintain a HD map

Autonomous Driving: on mobile SoC ?

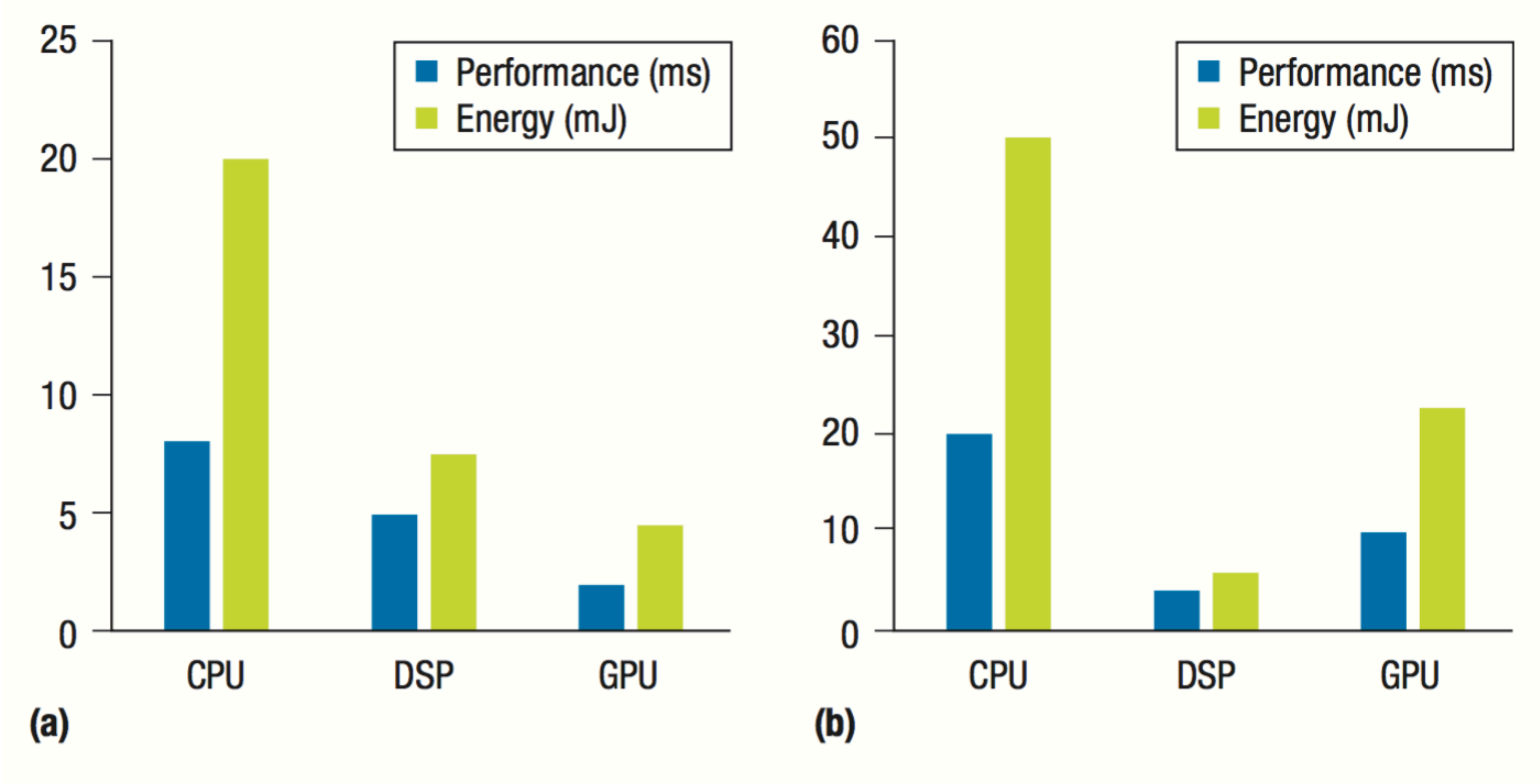
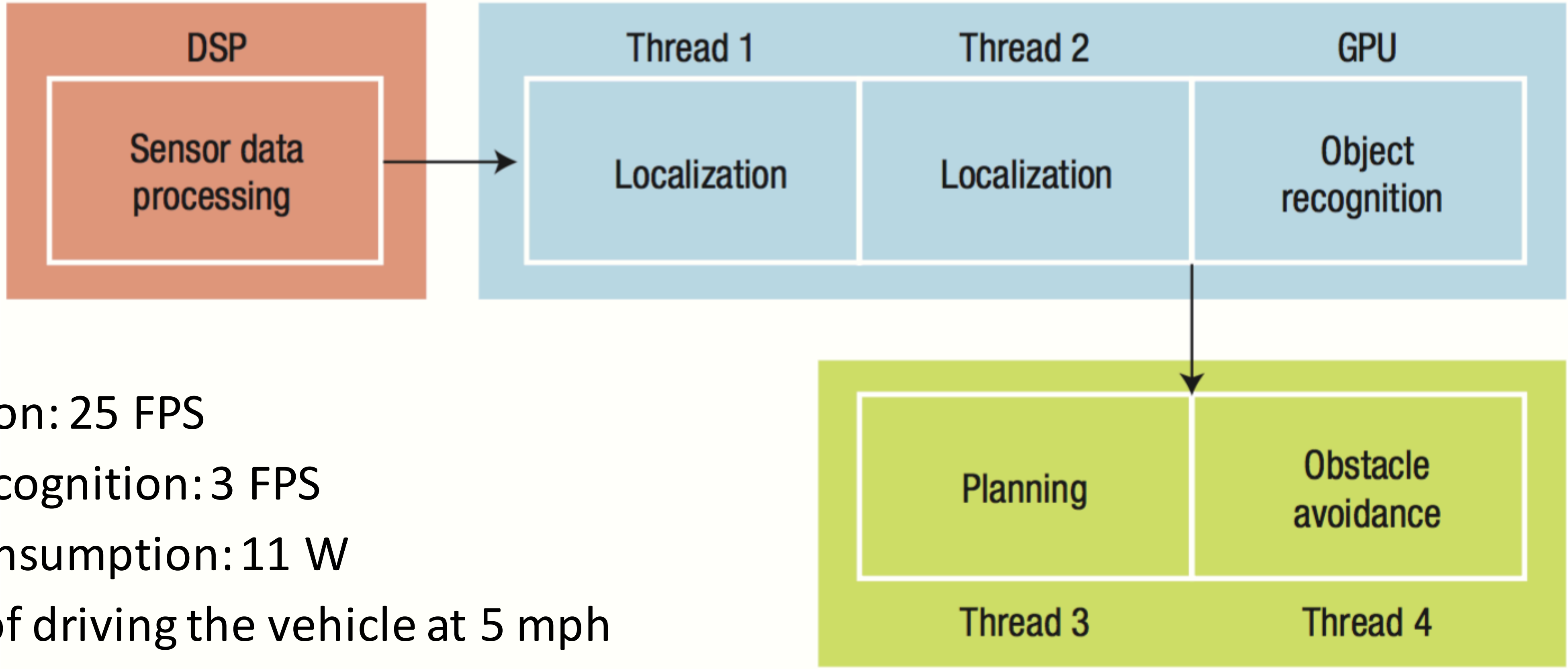


FIGURE 2. Performance and energy in (a) convolution and (b) feature-extraction tasks. In (a), the GPU takes only 2 ms and uses only 4.5 milliJoules (mJ) to complete convolution tasks. In (b), the digital signal processor (DSP) is the most efficient unit for feature extraction, taking 4 ms and consuming only 6 mJ to complete a task.

Mobile SoC:

- Quad-core CPU @ 2.2 GHz
- Hexagon 680 DSP
- Adreno 530 GPU
- Peak power ~ 15 W

Autonomous Driving: Heterogeneous Computing



- Localization: 25 FPS
- Object recognition: 3 FPS
- Power consumption: 11 W
- Capable of driving the vehicle at 5 mph

PERCEPTIN

普思英察

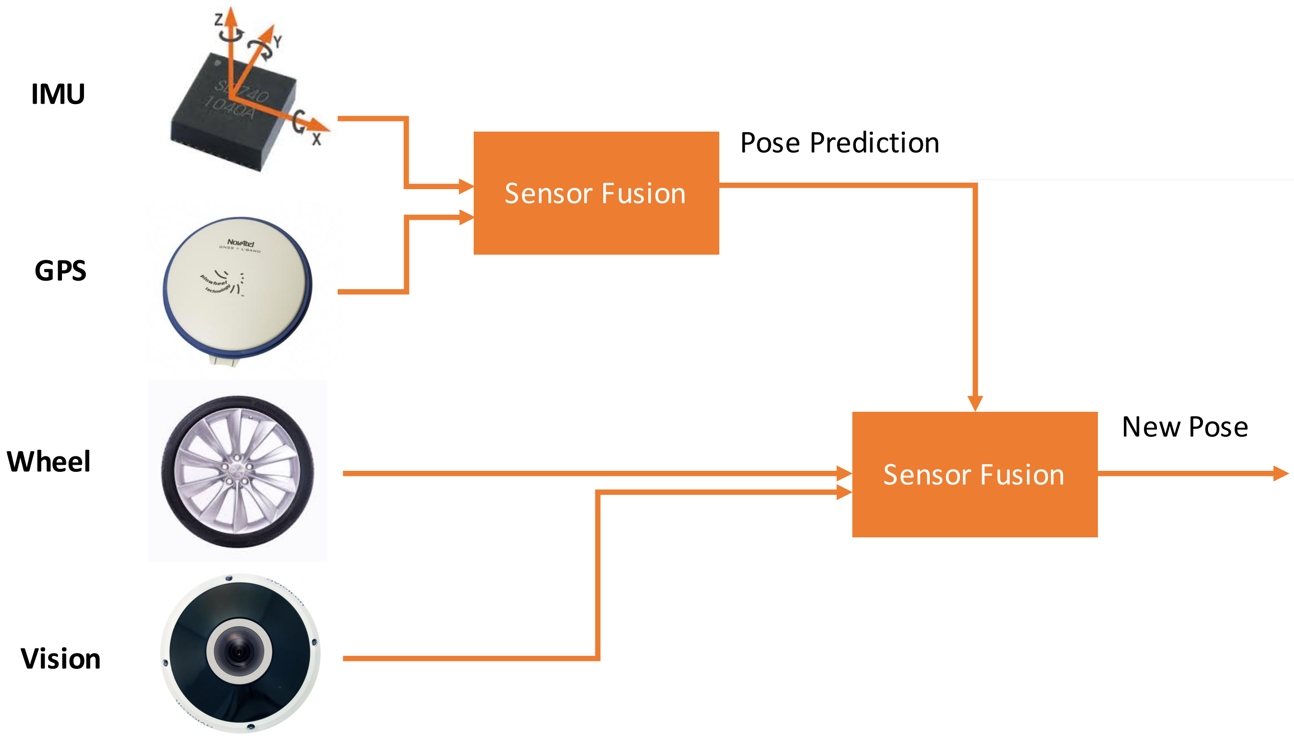
Computer Vision for Perception and Localization

- Four-way synchronized images: stereo 360-degree views
- Embedded with IMU and GPS, interface with wheel

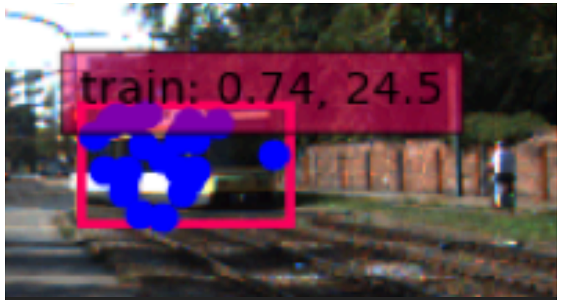
Localization

Perception

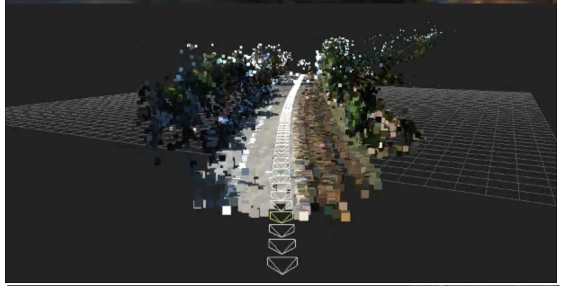
Planning and Control



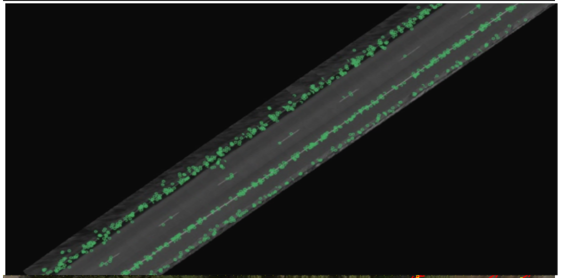
Universal High Precision Visual Map



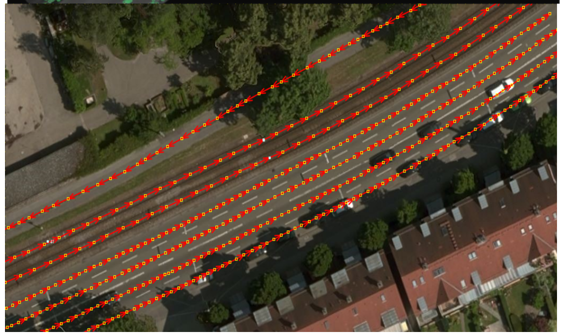
Layer 4: semantic Information



Layer 3: spatial features



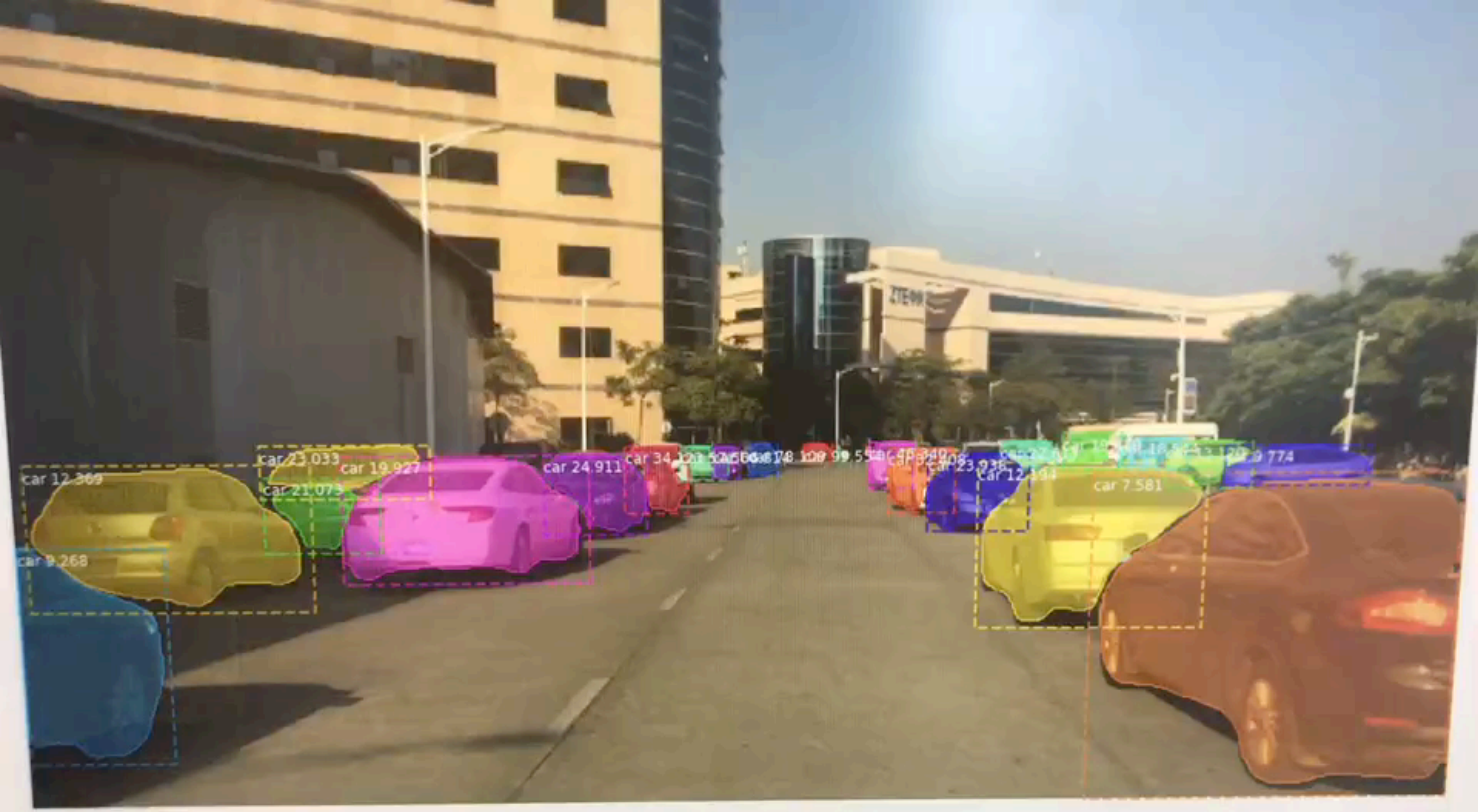
Layer 2: ground features



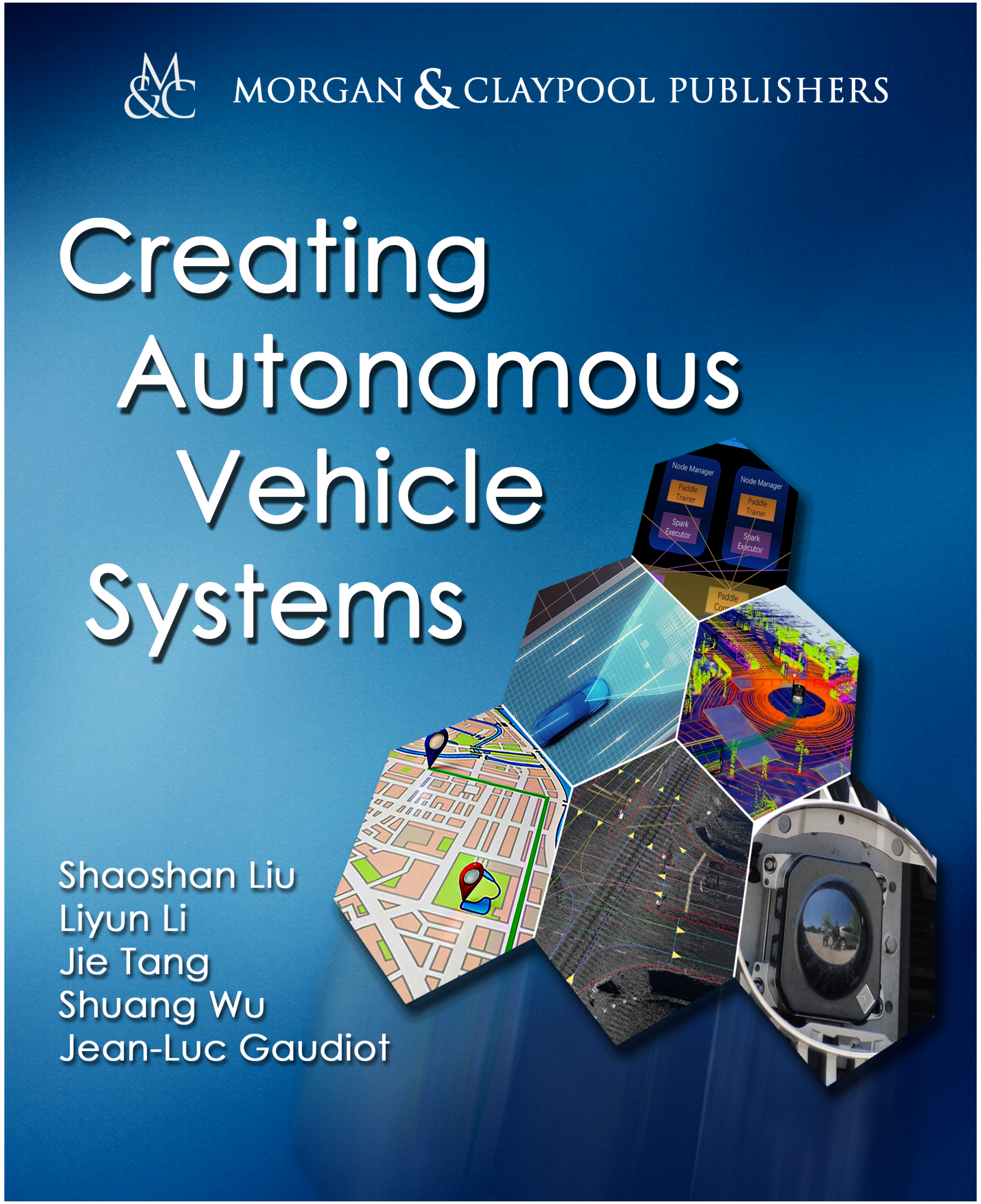
Layer 1: Digital map with lane-level annotation

PERCEPTIN

10,000 USD Autonomous Vehicle


car 23.033
car 12.369
car 19.927
car 21.073
car 24.911
car 9.268
car 12.194
car 7.581


PERCEPTIN



Teaching Autonomous Driving Using a Modular and Integrated Approach

Jie Tang[1], Shaoshan Liu[2], Songwen Pei[3], Stéphane Zuckerman[4], Chen Liu[5], Weisong Shi[6], and Jean-Luc Gaudiot[7]

Browse Journals & Magazines > Computer > Volume: 50 Issue: 8

Computer Architectures for Autonomous Driving

Computer

A Unified Cloud Platform for Autonomous Driving

Issue No. 12 - December (2017 vol. 50)
ISSN: 0018-9162
pp: 42-49
DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/MC.2017.4451224

Shaoshan Liu , PerceptIn
Jie Tang , South China University of Technology
Chao Wang , Baidu
Quan Wang , Baidu
Jean-Luc Gaudiot , University of California, Irvine