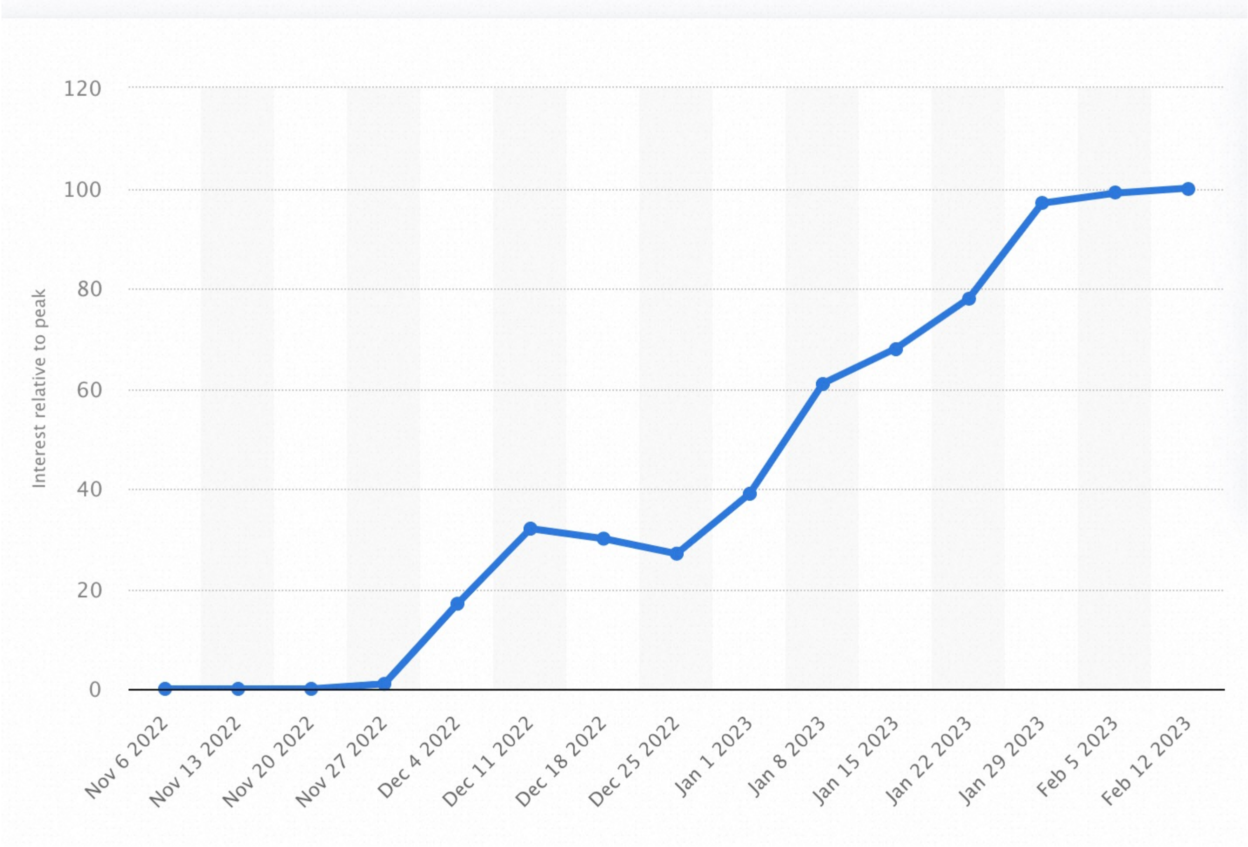
Autonomous Driving
Towards Reducing Human Efforts in Visual Perception and Beyond

Dr. Kaicheng Yu,
PI of Autonomous Intelligence Lab,
Westlake University

2024/04/08

Large AI Model Changes The World

ChatGPT is the **fastest app** reaches 1M Users
Only has **1** feature, Chat with GPT



Google Trends of ChatGPT

1. Statistica.com, https://www.statista.com/statistics/1366930/chatgpt-google-search-weekly-worldwide/, accessed on May 26th
2. Twitter Watcher.Guru, https://watcher.guru/news/how-long-did-it-take-chatgpt-to-reach-1-million-users, accessed on May 31th

Large AI Model Will Change The World Virtually

User

Computer (Desktop or Mobile)

Large AI Models

Closed Sourced

Alibaba - Tongyi

Baidu

ChatGPT + Bing

Google Bard

Claude

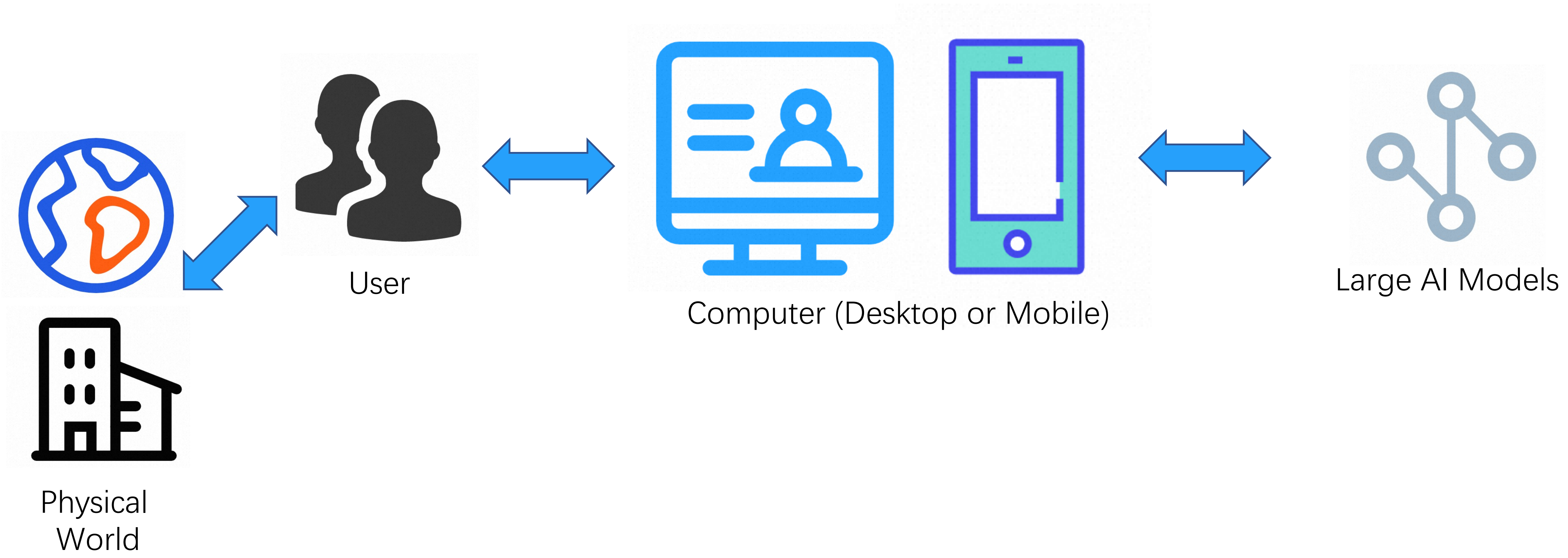
Open Sourced

Vicuna

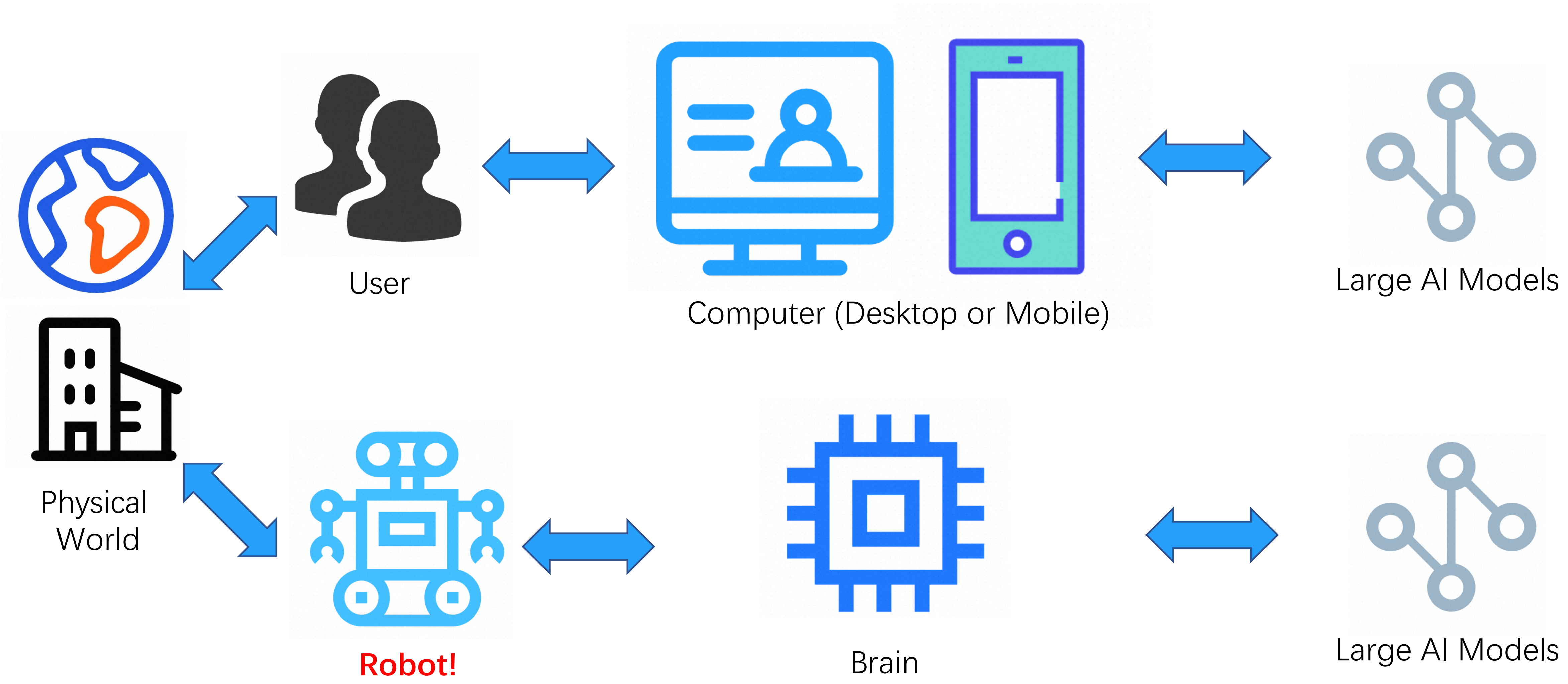
Generative Agents

AutoGPT

How does AI Model interact with physical world?



How does AI Model interact with physical world?



Autonomous Driving Vehicle Is Also A Robot

Autonomous Driving
Understand and Act in 3D World

Bus

Taxi

Heavy Truck

Carrier

Large-scale deployment of AV across China

Carrier
Largest Autonomous Driving in logistic

200+ Cities

800+ AutoVehicle

50M+ orders

Truck
Research -> Product

50+ routes across China

30+ test vehicles

100M+km test milage

Heavy Truck
Preliminary Exploration

Built 20+ Auto-Truck

Cainiao, Shentong

Release in 2027

PART I: General introduction of Autonomous Driving System (ADS)

Source: STDevCon19_7.5_Overview of ADAS-Active-Safety

Automotive ADAS Systems

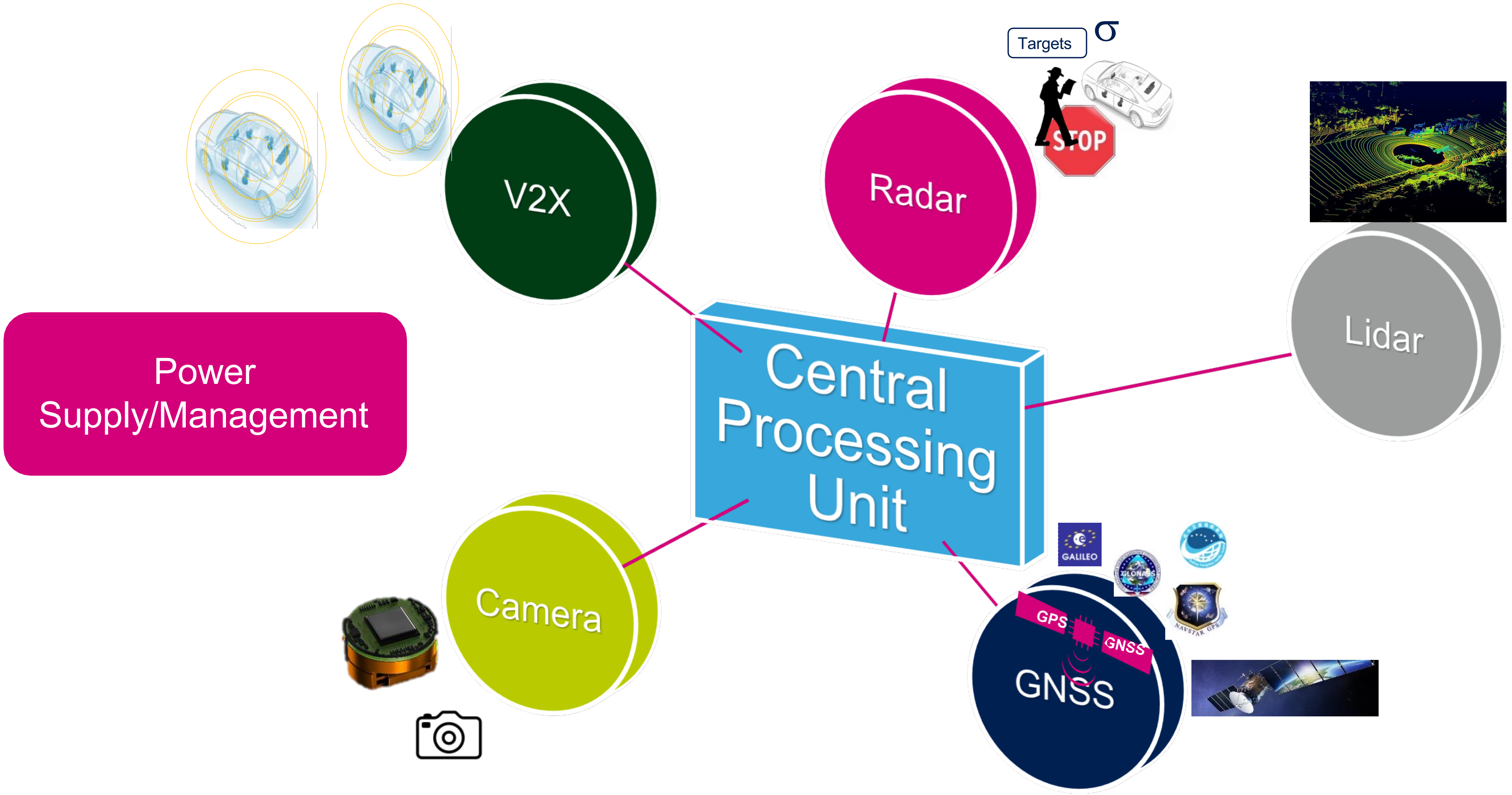
Overall Automotive ADAS System

Table of Contents

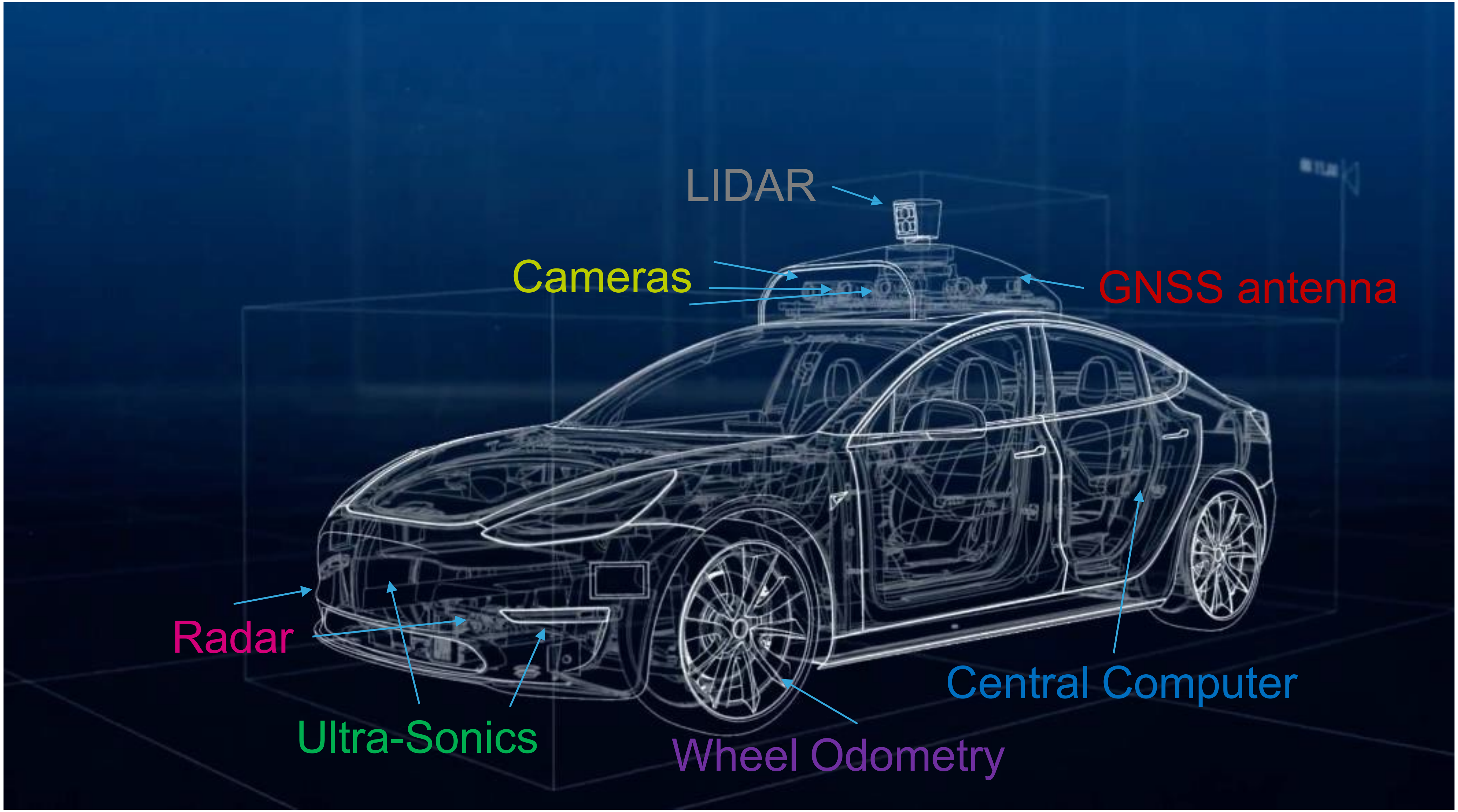
Automotive ADAS Systems

ADAS Overview

Overview of ADAS Technologies



ADAS Sensors - Needed for Perception



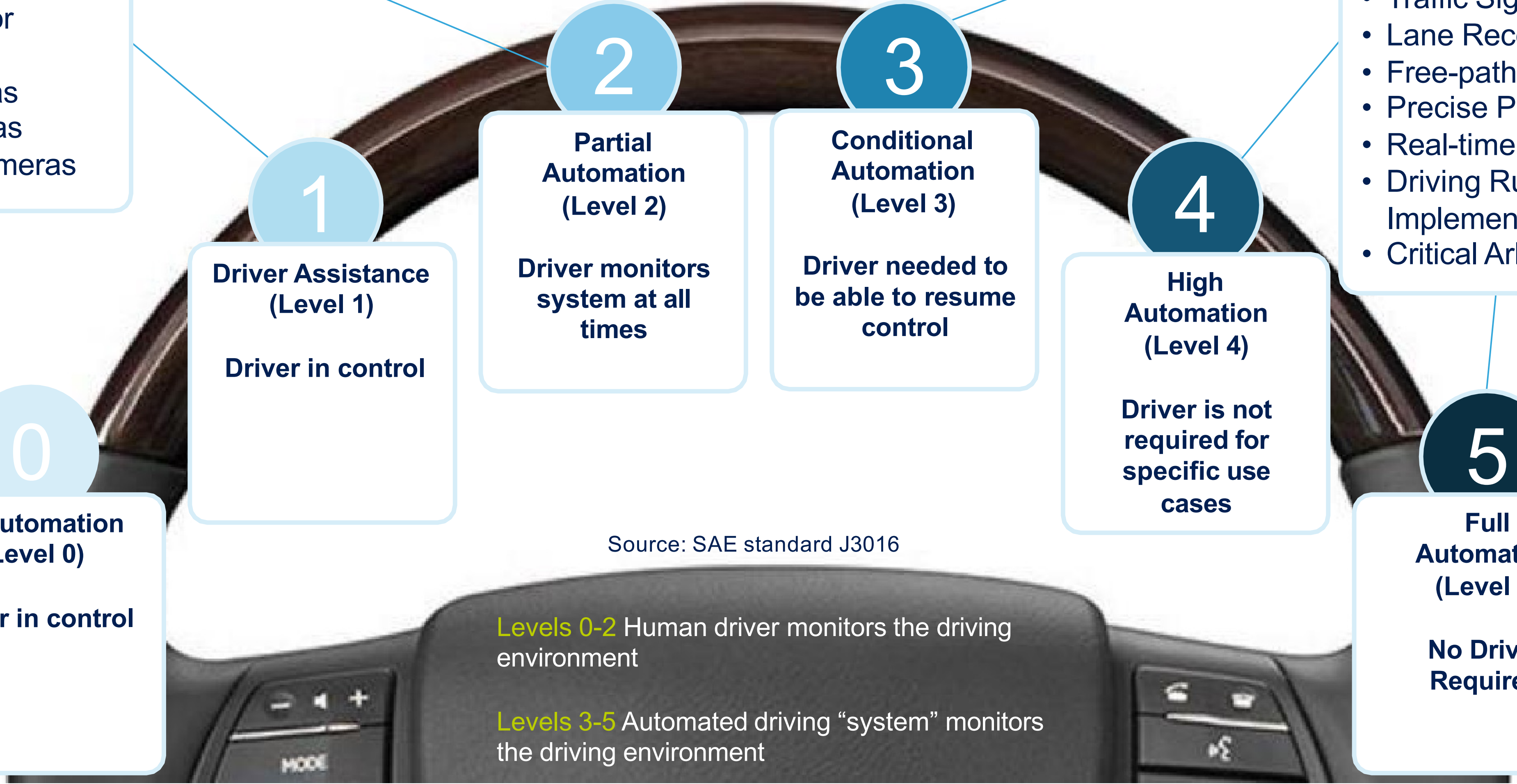
The 5 Levels of Vehicle Automation

Adding Senses

- Accelerometers and Gyro
- Steering Wheel Angle
- Ultrasonic sensors
- Front Radar Sensor
- Blind Spot sensor
- Rear View Cameras
- Front View Cameras
- Surround View Cameras

Learning to Drive

- Systems Networking
- Sensor Fusion
- Distance Measurement
- Traffic Sign Recognition
- Lane Reconstruction
- Free-path Definition
- Precise Positioning
- Real-time Mapping
- Driving Rules Implementation
- Critical Arbitration



0

No Automation (Level 0)

Driver in control

1

Driver Assistance (Level 1)

Driver in control

2

Partial Automation (Level 2)

Driver monitors system at all times

3

Conditional Automation (Level 3)

Driver needed to be able to resume control

4

High Automation (Level 4)

Driver is not required for specific use cases

5

Full Automation (Level 5)

No Driver Required

Source: SAE standard J3016

Levels 0-2 Human driver monitors the driving environment

Levels 3-5 Automated driving "system" monitors the driving environment

Sensor Fusion is Key to Autonomous

No sensor type works well for all tasks and in all conditions, so sensor fusion will be necessary to provide redundancy for autonomous functions

Most likely used fusion solution in future — Good — Fair — Poor

	Camera	Radar	LiDAR	Ultrasonic	LiDAR+Radar+ Camera
Object detection	Fair	Good	Good	Good	Good
Object classification	Good	Poor	Fair	Poor	Good
Distance estimation	Fair	Good	Good	Good	Good
Object edge precision	Good	Poor	Good	Good	Good
Lane tracking	Good	Poor	Poor	Poor	Good
Range of visibility	Fair	Good	Fair	Poor	Good
Functionality in bad weather	Poor	Good	Fair	Good	Good
Functionality in poor lighting	Fair	Good	Good	Good	Good

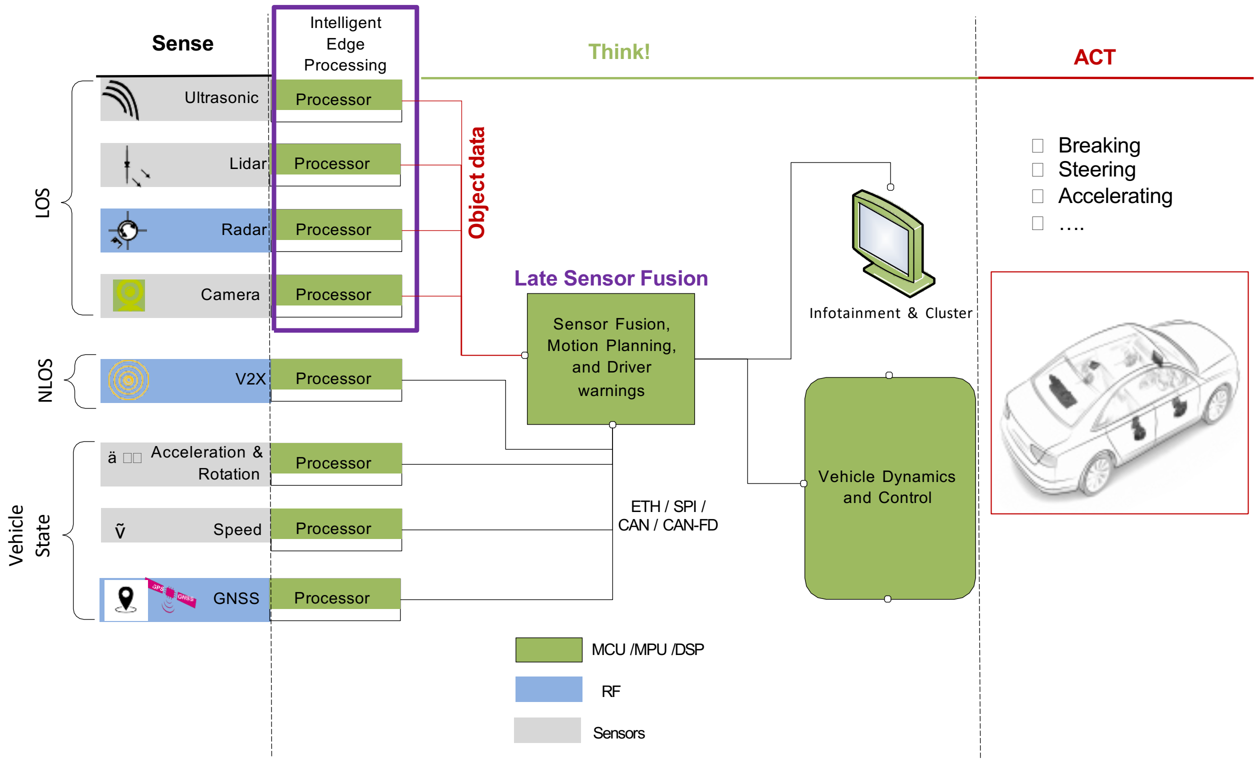
Source: Woodside Capital Partners (WCP), "Beyond the Headlights: ADAS and Autonomous Sensing", September 2016

Automotive ADAS Systems

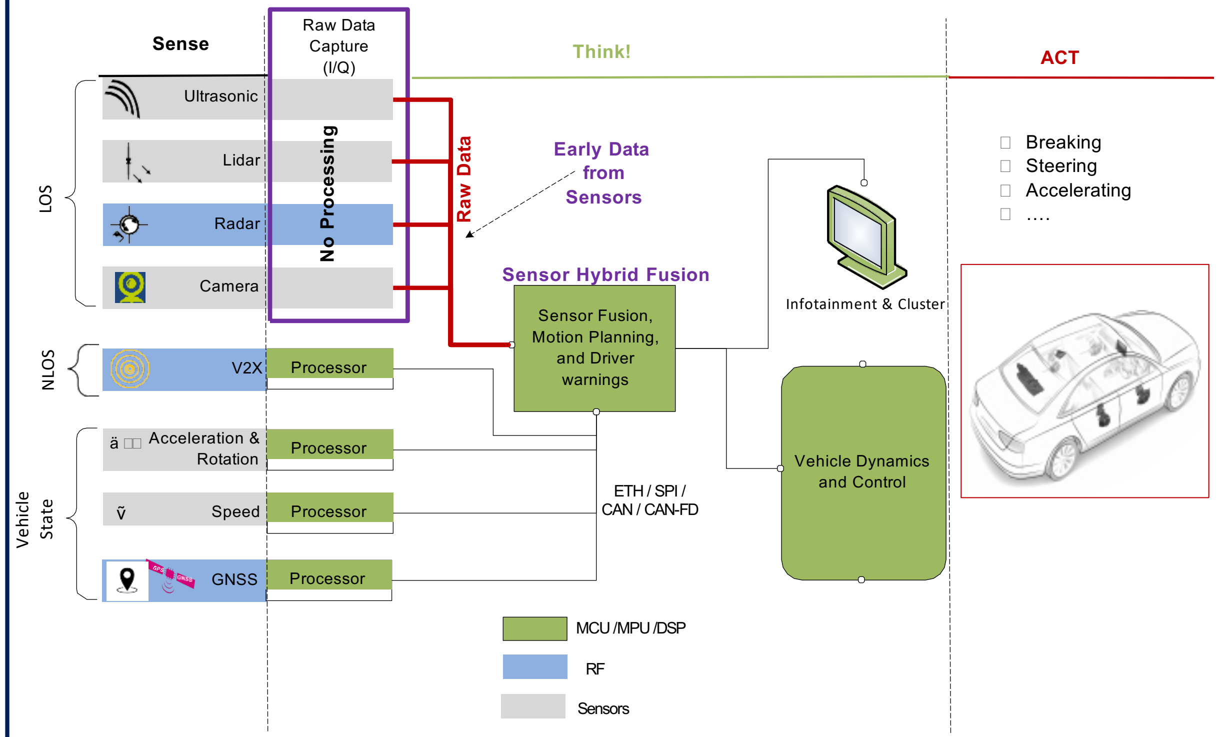
ADAS Vehicle Architectures

Distributed vs Centralized Processing

Distributed Processing with Object Level Fusion



Centralized Processing with Raw Data Fusion

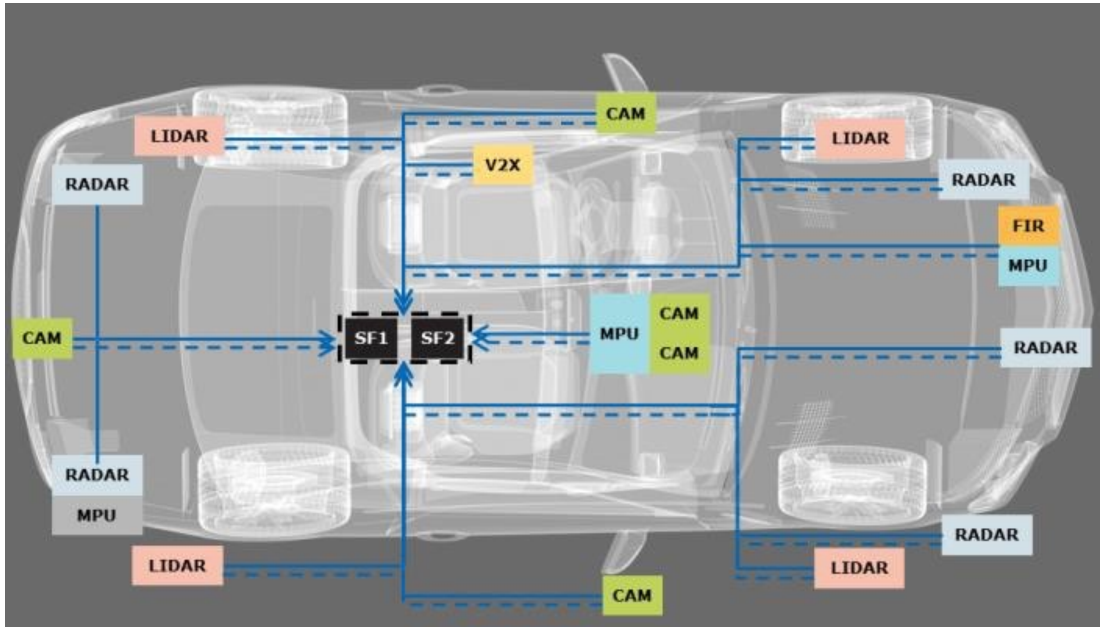


LOS: Line-of-Sight
NLOS: Non-Line-of-Sight

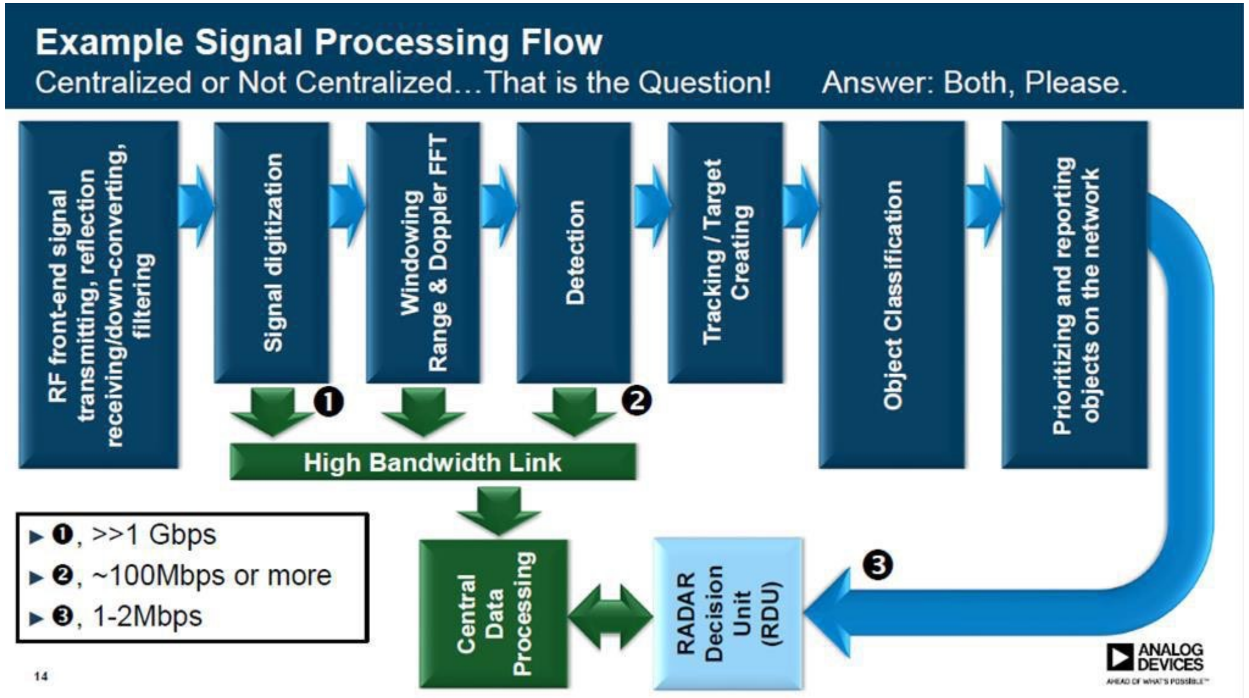
- Distributed Interfaces
  - ETH, SPI, I2C, CAN, CAN-FD
    - RADAR, Ultrasonic, V2X, IMU, Wheel Odomerty, GNSS
  - MIPI(CSI-2), GMSL(Maxim), FPD-Link(TI), PCIe, HDBaseT(Valens)
    - Video Cameras?
    - Lidar?

- Centralized Interfaces
  - ETH, SPI, I2C, CAN, CAN-FD
    - V2X, IMU, Wheel Odomerty, GNSS
  - MIPI(CSI-2), GMSL(Maxim), FPD-Link(TI), PCIe, HDBaseT(Valens)
    - Radar, Ultrasonic
    - Cameras
    - Lidar?

Distributed vs Centralized Processing



Source: 2018 IHS Markit – "Autonomous Driving-The Changes to come"



Source: ADI

- What are the Data rates requirements for each sensor?
  - Centralized (i.e. SERDES?) vs Distributed (i.e. ETH?)

- Example: 4-5 Corner Radars are utilized in high end/premium vehicles.

Automotive ADAS Systems

Vision (Cameras) System

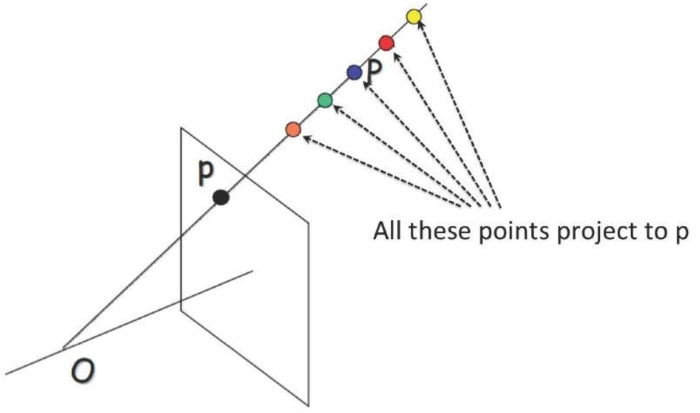
Camera

- Essential for correctly perceiving environment
- Richest source of raw data about the scene - only sensor that can reflect the true complexity of the scene.
- The lowest cost sensor as of today
- Comparison metrics:
  - Resolution
  - Field of view (FOV)
  - Dynamic range
- Trade-off between resolution and FOV?

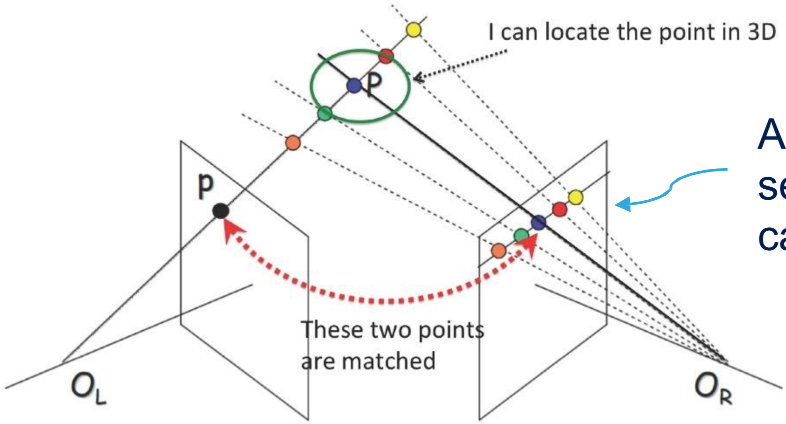
Camera-Stereo

- Enables depth estimation from image data

All points on projective line to P map to p

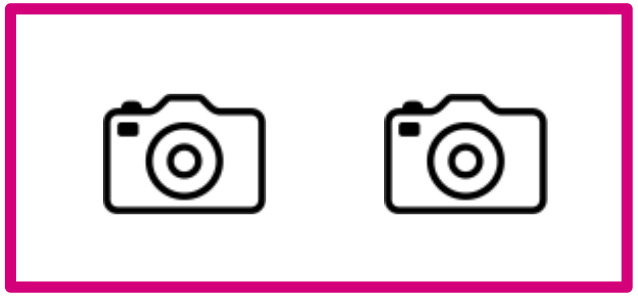


One camera



Add a second camera

Find a point in 3D by triangulation!

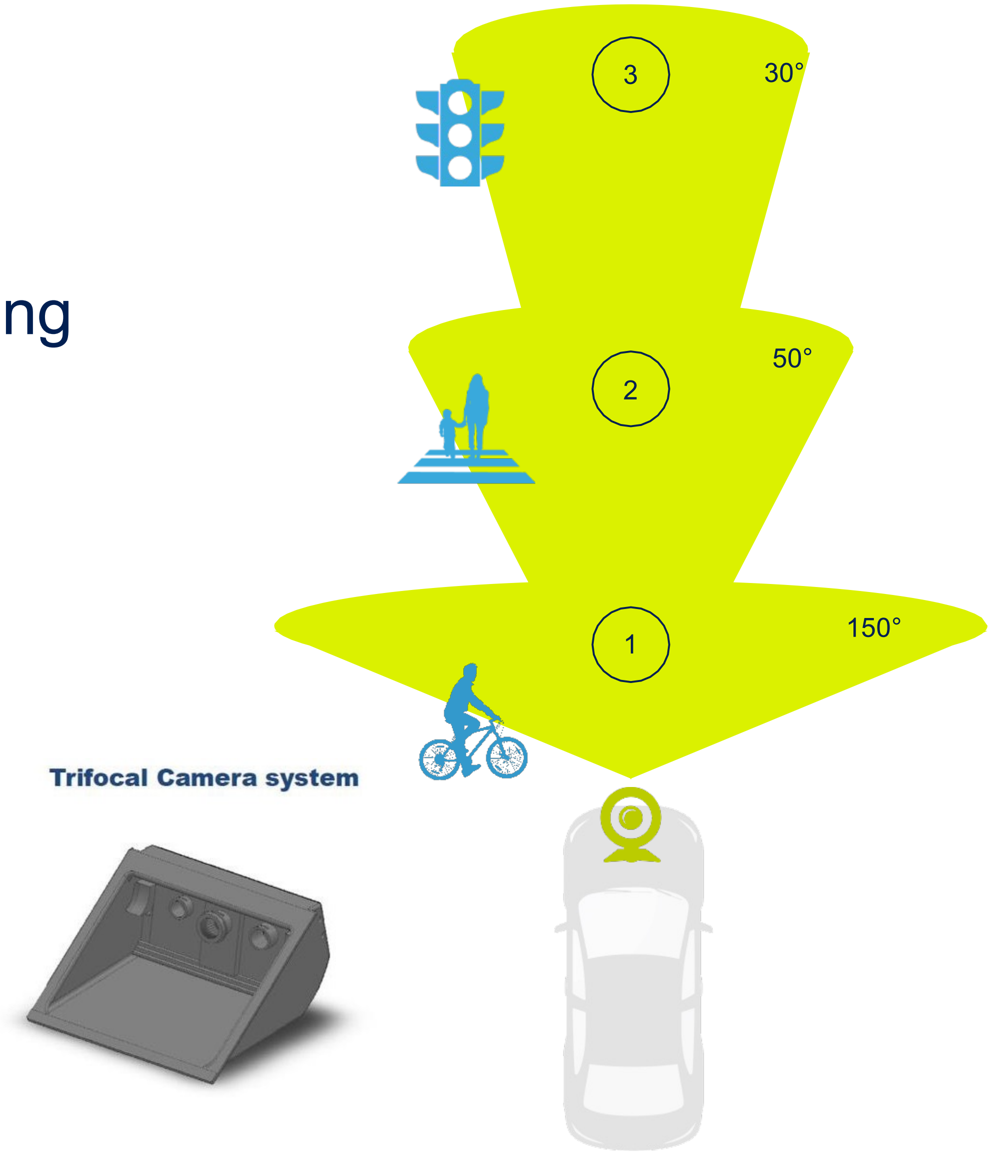


Left and right images

Source: Sanja Fidler, CSC420: Intro to Image Understanding

The Next Phase for Vision Technology

- From sensing to comprehensive perception
- Machine learning used already for object sensing
- Autonomous driving needs
  - Path planning based on holistic cues
  - Dynamic following of the drivable area
- Deep learning is now being applied



Machine Vision: ST & Mobileye

EyeQ3™ 3rd Generation vision processor

- Detection of driving lanes
- Recognition of traffic signs
- Detection of pedestrians and cyclists
- Seeing obstacles how the human eye sees them
- Adapting cruise speed
- Emergency braking when car ahead slows suddenly

Partnership

EyeQ4™ 4th Generation enables

- Detection of more objects, more precisely
- More features required for automated driving Free-space Estimation, Road Profile Reconstruction
- Monitoring of environmental elements (fog, ice, rain) and their safety impact
- Detailed understanding of the road conditions allowing automatic suspension and steering adjustment
- Highly automated vehicles

EyeQ5™
EyeQ5™

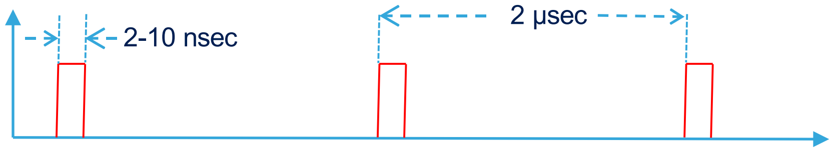
The Road to Full Autonomous Driving: Mobileye and ST to Develop EyeQ®5 SoC targeting Sensor Fusion Central Computer for Autonomous Vehicles

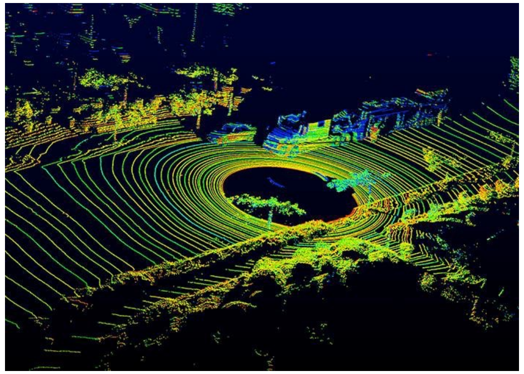
LiDAR Technology Overview

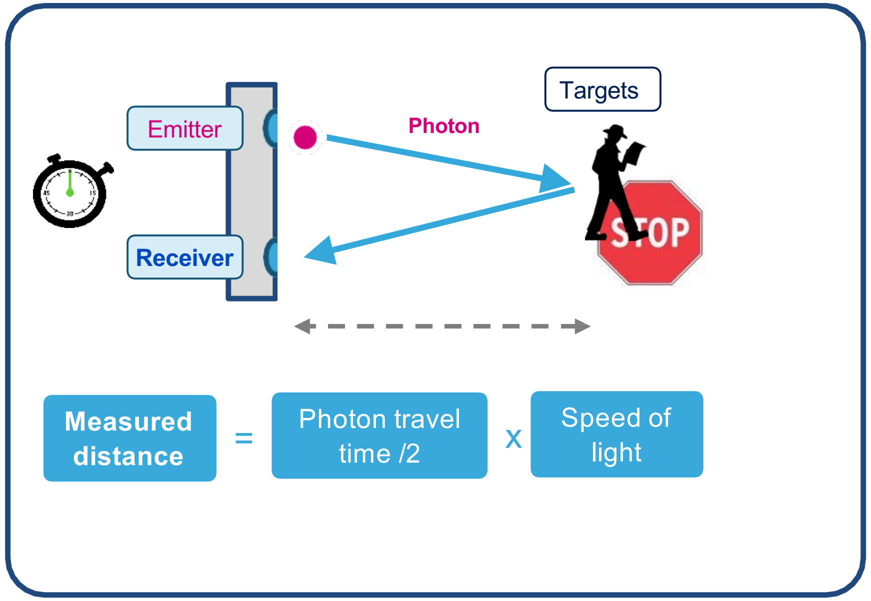
- LiDAR (light detecting and ranging, or “light radar”) sensors send one or more laser beams at a high frequency and use the Time-of-Flight principle to measure distances. LiDAR capture a high-resolution point cloud of the environment.

- Can be used for object detection, as well as mapping an environment
  - Detailed 3D scene geometry from LIDAR point cloud

- LiDAR uses the same principal as ToF sensor, but at much longer distances, minimum 75M for “near field” and 150-200M for “far field”.







Automotive ADAS Systems

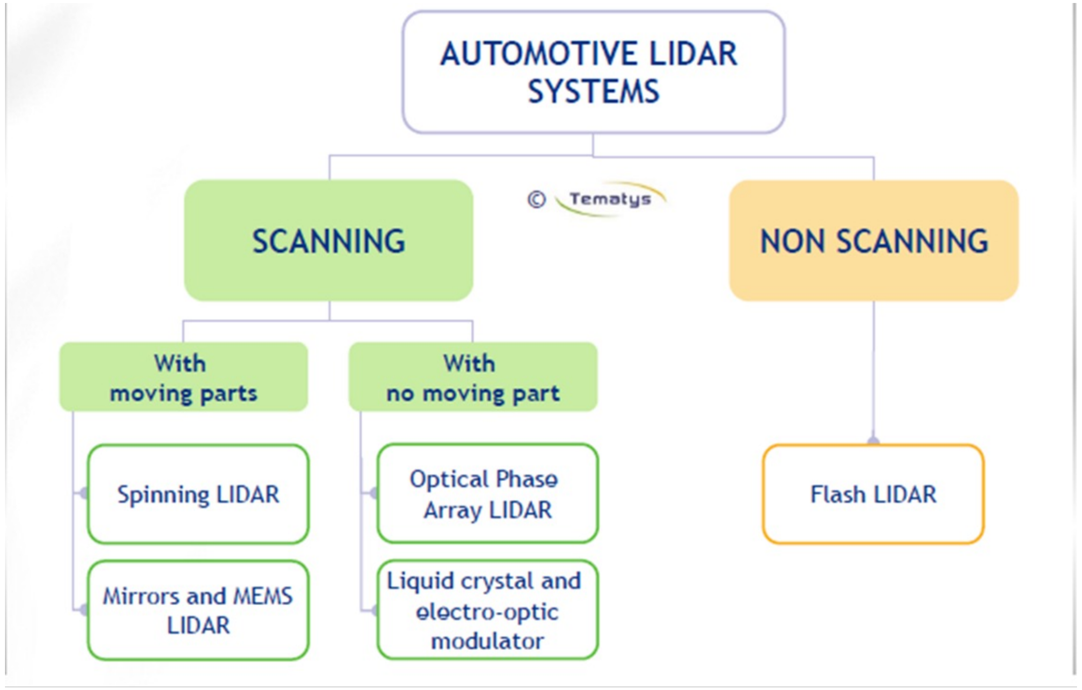
LiDAR System

LiDAR Techniques

- There are multiple techniques currently under evaluation for LiDAR including rotating assembly, rotating mirrors, Flash (single Tx source, array Rx), scanning MEMS micro-mirrors, optical phased array.

- From a transmitter/receiver (Tx/Rx) perspective the following technologies need to be developed or industrialized for automotive.
  - MEMS Scanning Micro-mirror technologies
  - SPAD (Single Photon Avalanche Detectors) - Rx
  - 3D SPAD - Rx
  - Smart GaN (Gallium nitride)

- Comparison metrics:
  - Number of beams: *8,16, 32, and 64 being common sizes*
  - Points per second: *The faster, the more detailed the 3D point cloud can be*
  - Rotation rate: *higher rate, the faster the 3D point clouds are updated*
  - Detection Range: *dictated by the power output of the light source*
  - Field of view: *angular extent visible to the LIDAR sensor*



Source: J. Cochard et.al., "LiDAR Technologies for the Automotive Industry", Tematsys, June 2018

Upcoming: Solid state LIDAR!

LiDAR Summary

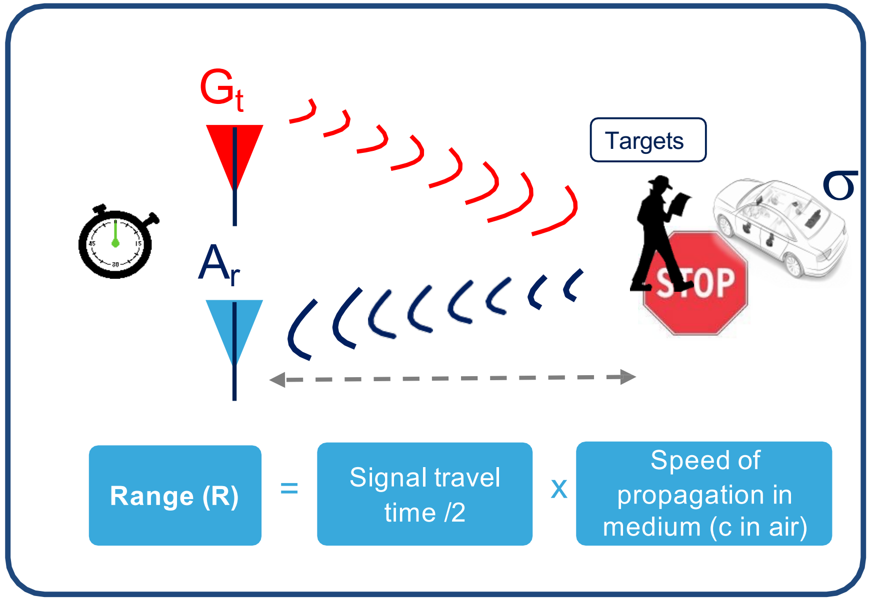
- Autonomous vehicles have been around for quite some time but only now the technologies are available for practical implementations
- No single sensor solution exists to cover all aspects – range, accuracy, environmental conditions, color discrimination, latency etc.
  - Multi-sensor fusion and integration will be a must
  - Each technology attempts to solve the overall problem while having multiple limitations
- Many LiDAR solutions (technologies) are available or being proposed with no clear winners
- Market is still in very early stage of development and experimentation
- When and which technology or system will be widely adopted and mass production starts is still unknown

Automotive ADAS Systems

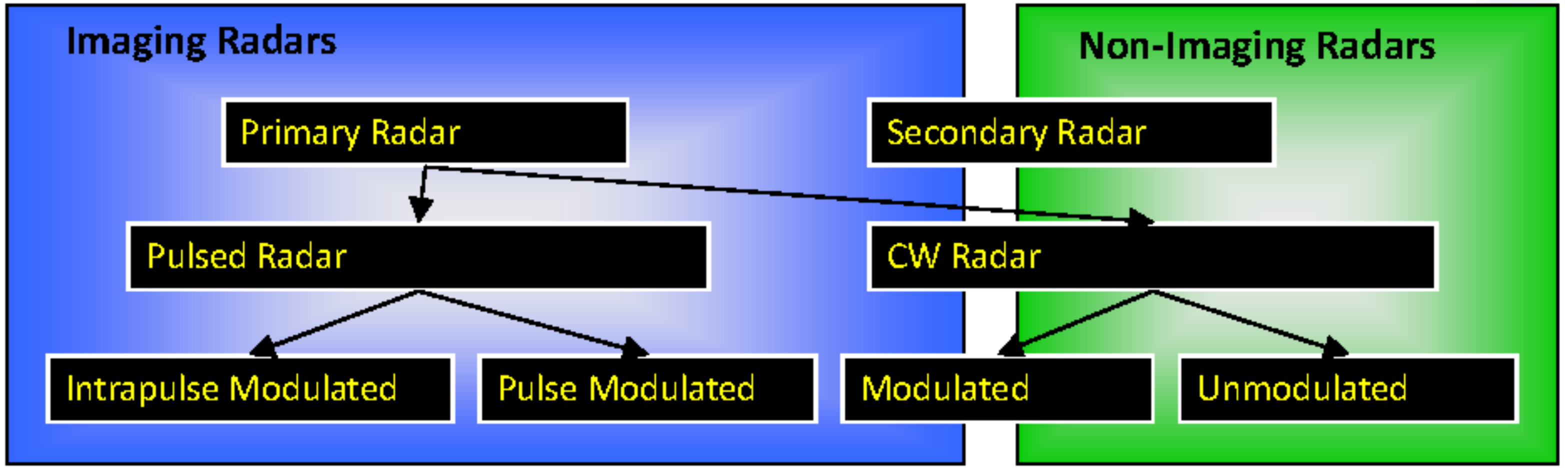
Radar Systems

RADAR Technology Overview

- RADAR (**RA**dio **D**etection and **R**anging) is one necessary sensor for ADAS (Advanced Driver Assistance System) systems for the detection and location of objects in the presence of interference; i.e., noise, clutter, and jamming.
- Robust Object Detection and Relative Speed Estimation
- Transmit a radio signal toward a target, Receive the reflected signal energy from target
- The radio signal can the form of “Pulsed” or “Continuous Wave”
- Works in poor visibility like fog and precipitation!
- Automotive radars utilize Linear FM signal, Frequency Modulated Continuous Wave (FMCW)
  - FM results in a shift between the TX and RX signals that allows for the determination of time delay, Range and velocity.



RADAR Techniques



- Definitions:
  - **Imaging Radar:** Forms a picture of the object or area
  - **Non-Imaging Radar:** Measures scattering properties of the object or area
  - **Primary Radar:** Transmits signals that are reflected and received
  - **Secondary Radar:** Transponder that responds to interrogation with additional info
  - **Pulsed Radar:** High power signals are only present for a short duration and repeated at specific intervals
  - **CW Radar:** Signal is present continuously

2013 Defence & Security Forum, EuMW

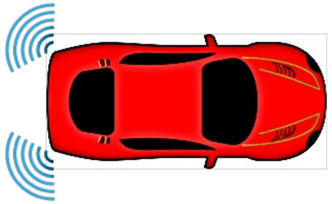
- Comparison metrics:
  - Range
  - Field of view
  - Position and speed accuracy
- Configurations:
  - Wide-FOV: Short Range
  - Narrow-FOV: Long Range

Source: Strategy Analytics Lunch & Learn the Market Session European Microwave Week 2013

Source: Rodhe & Schwarz - Automotive radar technology, market and test requirements, White paper – Oct 2018 (Salvo S. presentation)

Automotive Radar Vs. Automation Levels

< 2014
Level 1
Driver Assistance



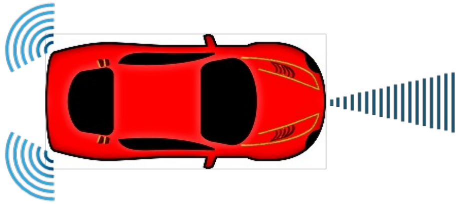
Object detection

2x SRR

Applications

BSD, LCA

2016
Level 2
Partial Automation



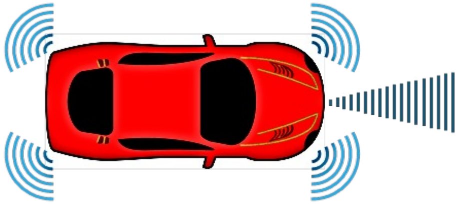
Object detection

2x SRR
1x LRR

Applications

BSD, RCW, LCA
ACC, AEB

2018
Level 3
Conditional Automation



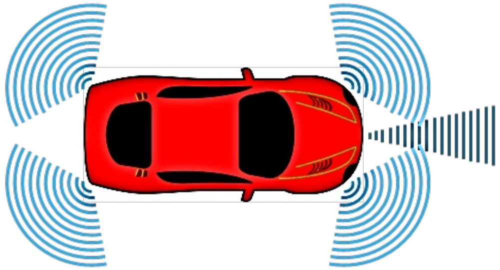
High resolution target separation

4x SRR
1x LRR

Applications

BSD, RCW, LCA
FCW, RCTA
ACC, AEB

2019 / 2020
Level 4
High Automation



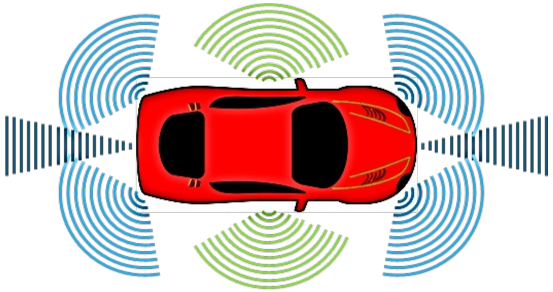
3D detection

4x SRR-MRR
1x LRR

Applications

BSD, LCA, RCTA
AEB pedestrian
ACC, AEB

> 2028
Level 5
Full Automation



360° object recognition

2x USRR
4x SRR-MRR
2x LRR

Applications

AVP, PA
BSD, LCA, RCTA
AEB pedestrian
ACC, AEB

USRR - Ultra Short Range Radar
SRR - Short Range Radar
MRR - Medium Range Radar
LRR - Long Range Radar

BSD - Blind Sport Detection
LCA - Lane Change Assist
RCW - Rear Collision Warning

ACC - Adaptive Cruise Control
AEB - Automatic Emergency Breaking
FCW - Forward Collision Warning

RCTA - Rear Cross Traffic Alert
AVP - Automated Valet Parking
PA - Parking Assist

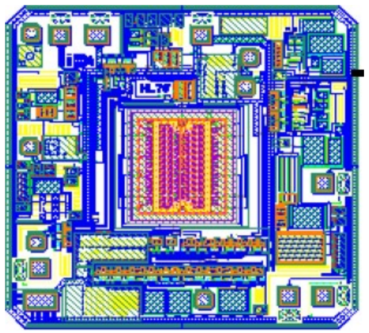
Automotive ADAS Systems

GNSS/IMU System

GNSS/IMU Positioning

- Global Navigation Satellite Systems and Inertial Measurement Units
- Direct measure of vehicle states
  - Positioning, velocity, and time (GNSS)
    - Varying accuracies: Real-time Kinematic (RTK-short base line), Precise Point Positioning (PPP), Differential Global Positioning System (DGPS), Satellite-based augmentation system (SBAS-Ionospheric delay correction)
  - Angular rotation rate (IMU)
  - Acceleration (IMU)
  - Heading (IMU, GPS)

GNSS/IMU



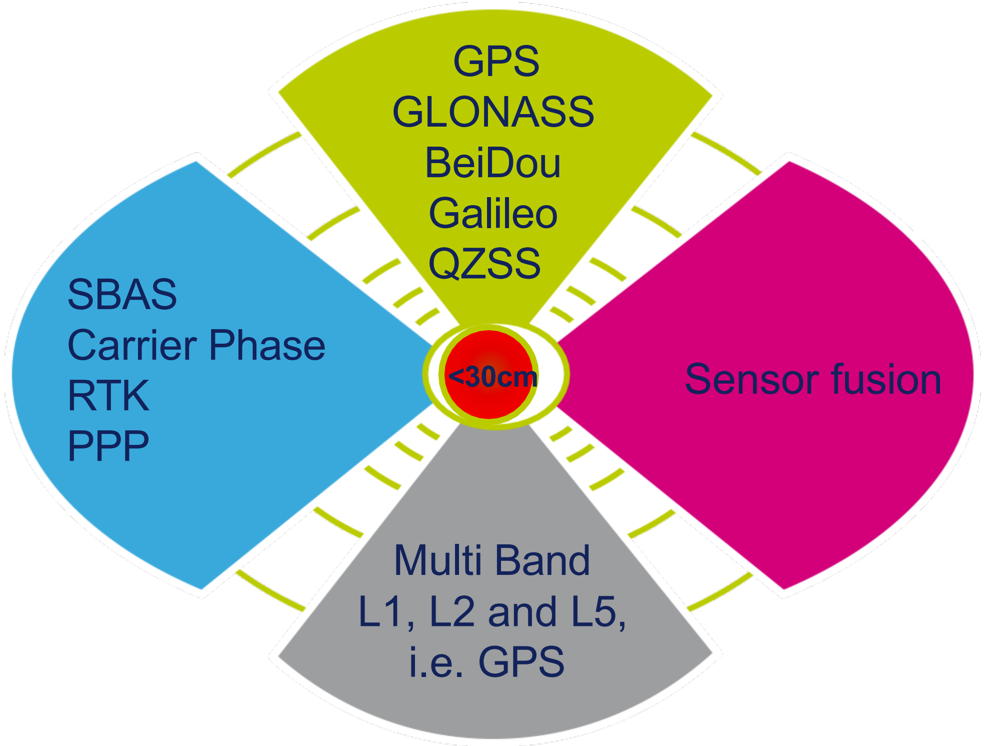
GNSS/IMU Positioning

More Precision Enables More Safety Features

Precise Positioning: Towards Autonomous Driving

Precise Positioning to enable < 30cm precision

- Lane detection
- Positioning data for V2X sharing
- Collision avoidance
- Autonomous parking
- Autonomous driving
- eCall accident location

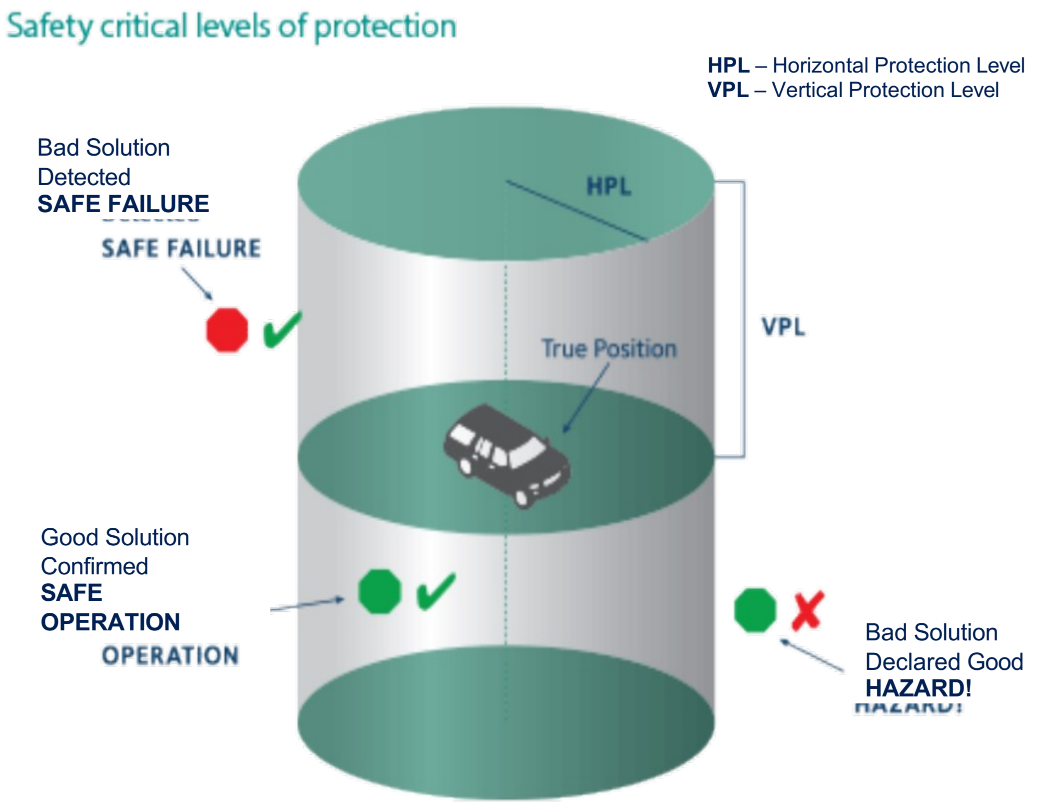


Precise GNSS is a Critical ADAS Sensor

Higher integrity requirements across safety-critical applications

- Semi- and Autonomous driving safety-related applications requirements **increase**
  - Higher safety levels
  - Added redundancy
  - More Robustness & integrity
  - Security

- **Teseo APP** (ASIL Precise Positioning) GNSS receiver, **new sensor** based on **ISO26262** concept with unique **Absolute and Safe** positioning information complementing **relative** positioning other sensor inputs(i.e. LIDAR, RADAR, etc.)



Courtesy of Hexagon PI

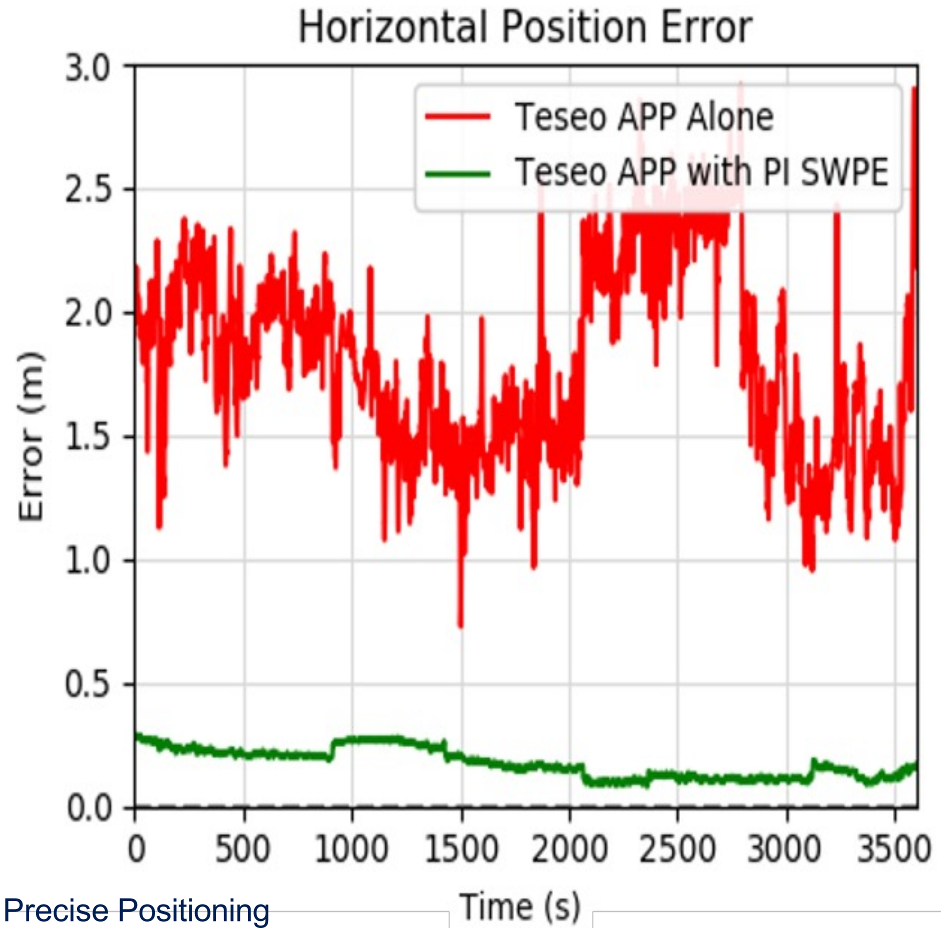
ST's GNSS Receiver Family for ADAS and AD

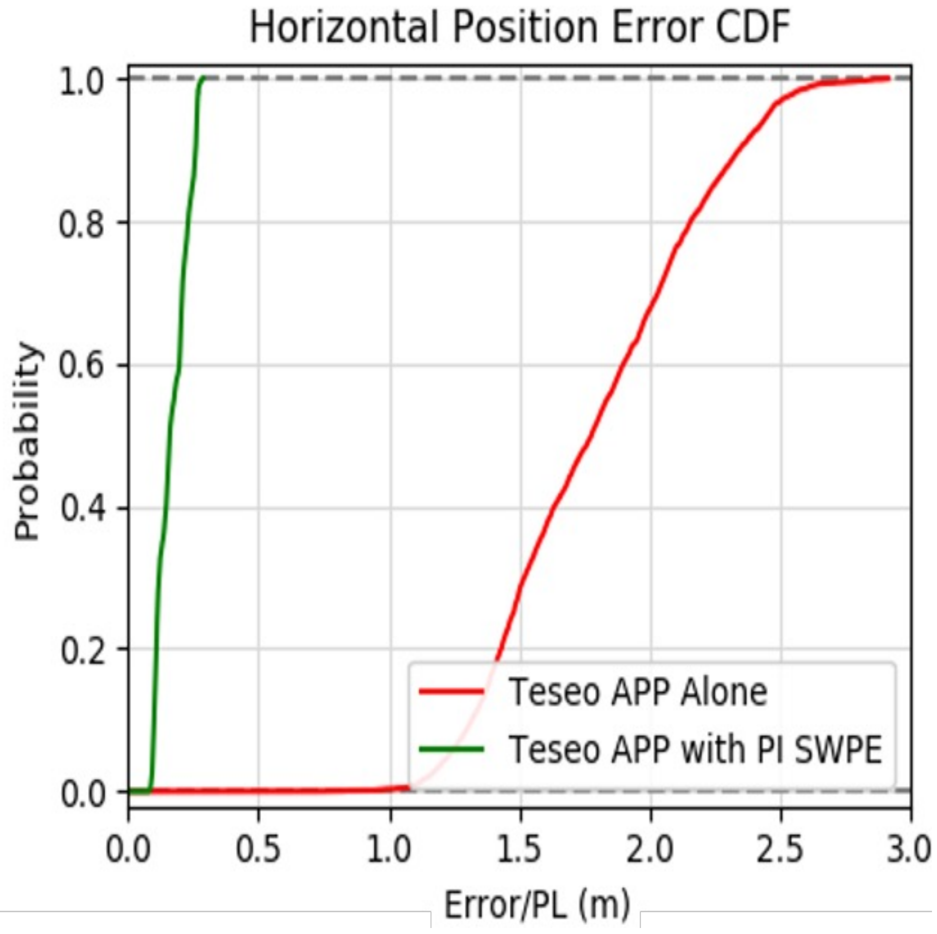
Precise GNSS is a Critical ADAS Sensor

GNSS Accuracy in Automotive Environment (using PPP – Precise Point Positioning)

Single Frequency (i.e. L1) multi-constellation/code-phase(1msec modulation signal)

Multi Frequency (i.e. L1, L2) multi-constellation/carrier-phase

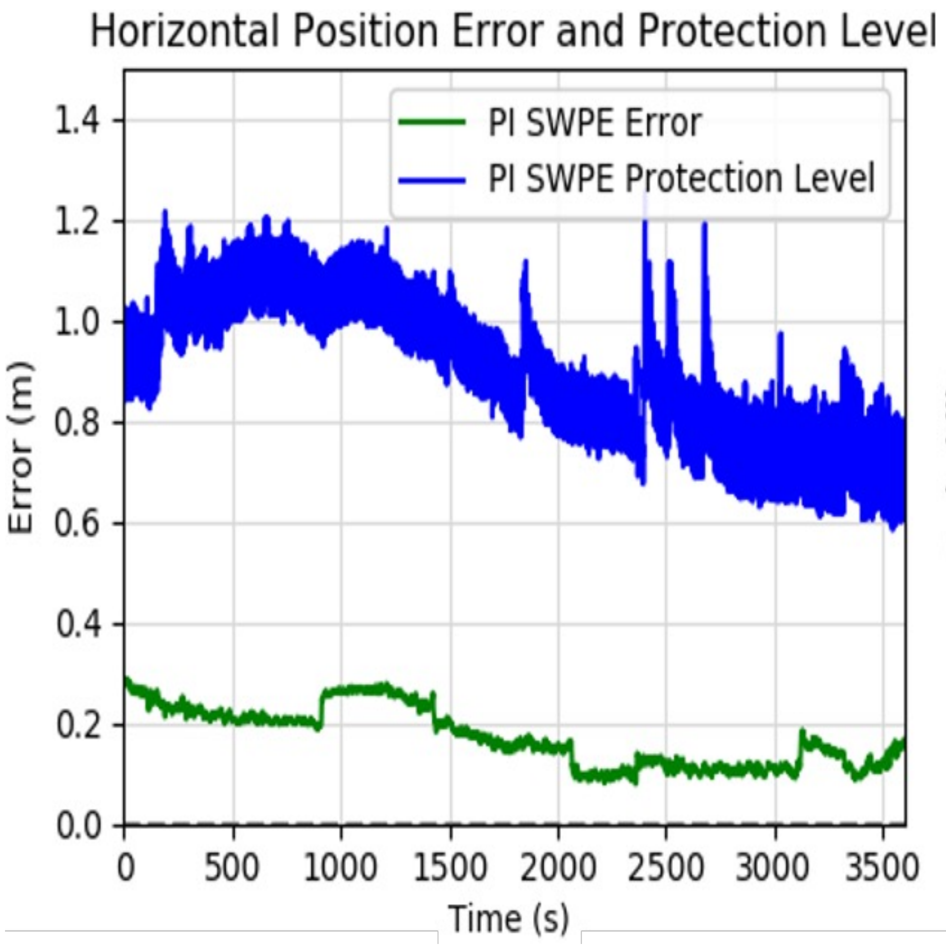


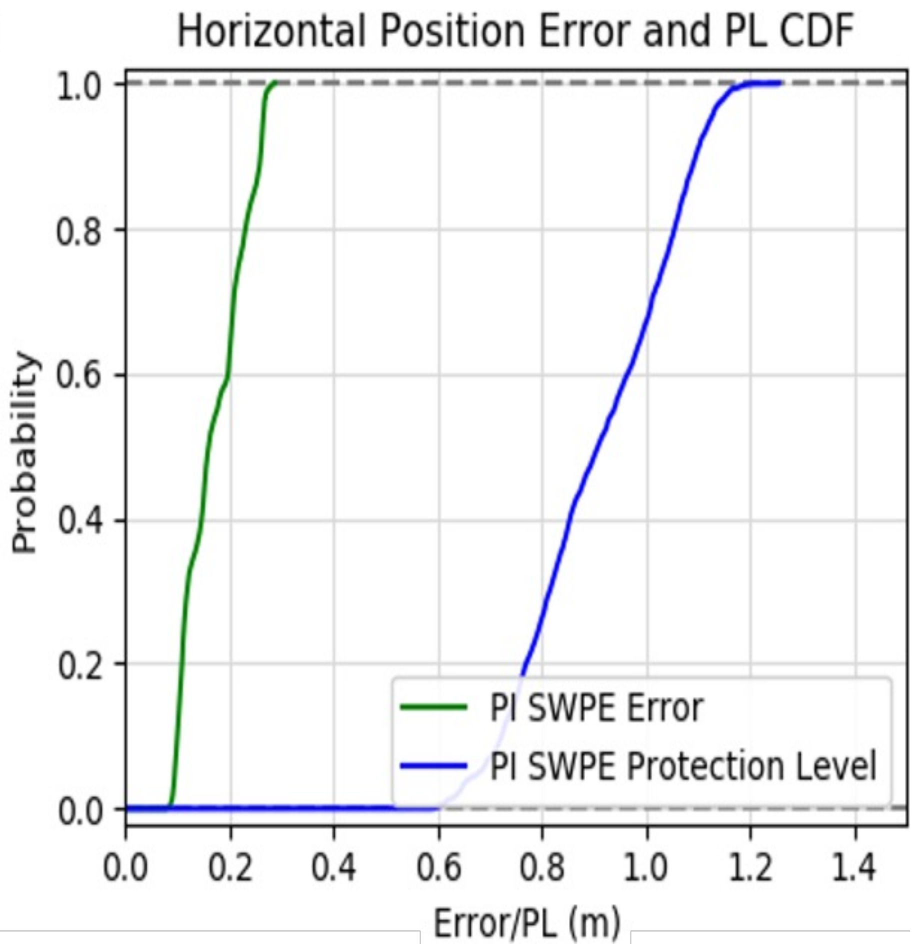


APP: ASIL Precise Positioning
SWPE: Software Positioning Engine

Precise GNSS is a Critical ADAS Sensor

GNSS Integrity – Protection Levels





Automotive ADAS Systems

V2X System

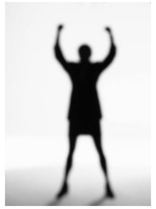
Vehicle-to-Everything (V2X)

V2X

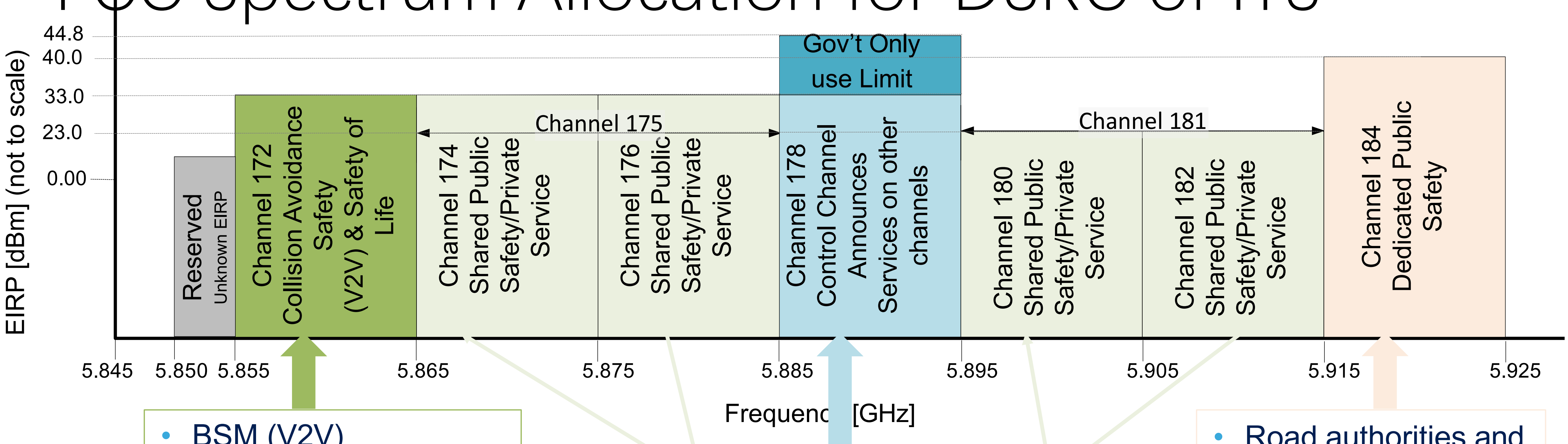
- **V2V** Vehicle-to-Vehicle
- **V2I** Vehicle-to-Infrastructure
- **V2M** Vehicle-to-Motorcycle
- **V2D** Vehicle-to-Device/object
- **V2P** Vehicle-to-Pedestrian







FCC Spectrum Allocation for DSRC of ITS



- BSM (V2V)
- MAP Message (V2I)
- SPAT (V2I)
- TX Power +20dBm

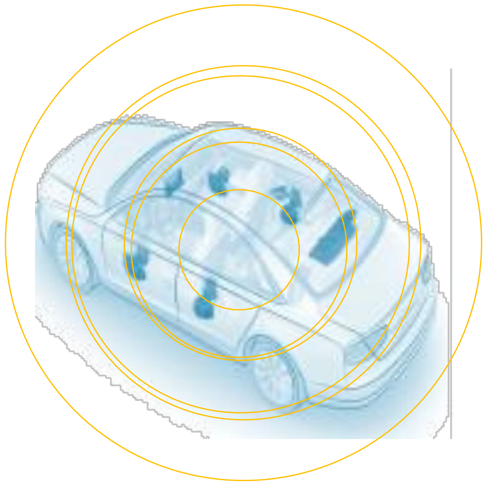
- Control Channel, Advertises and indicates how to access services on other "Service channels"

- Road authorities and public agencies primarily responsible for usage

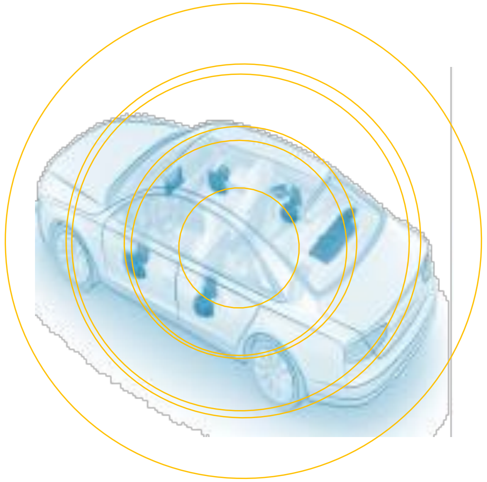
EIRP: Effective Isotropic Radiated Power
ITS: Intelligent Transportation Systems

Source: **Federal Communications Commission FCC 03-324**

DSRC

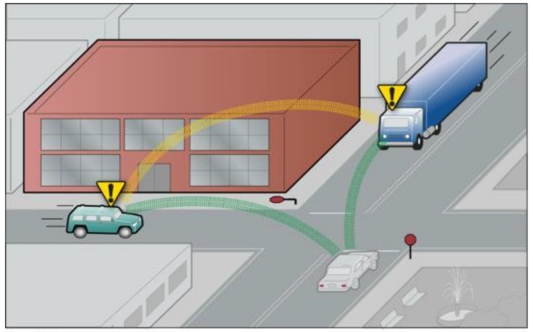


- Wireless Access in Vehicular Environments (WAVE)
  - Amendment to IEEE 802.11-2012 to support WAVE/DSRC
  - no authentication, no access point/no association
  - 5.8 – 5.9 GHz OFDM



- Fast Network Acquisition & low latency (<50msec)
- Priority for Safety Applications
- Interoperability
- Security and Privacy (ensured through a root certification system)

NLOS

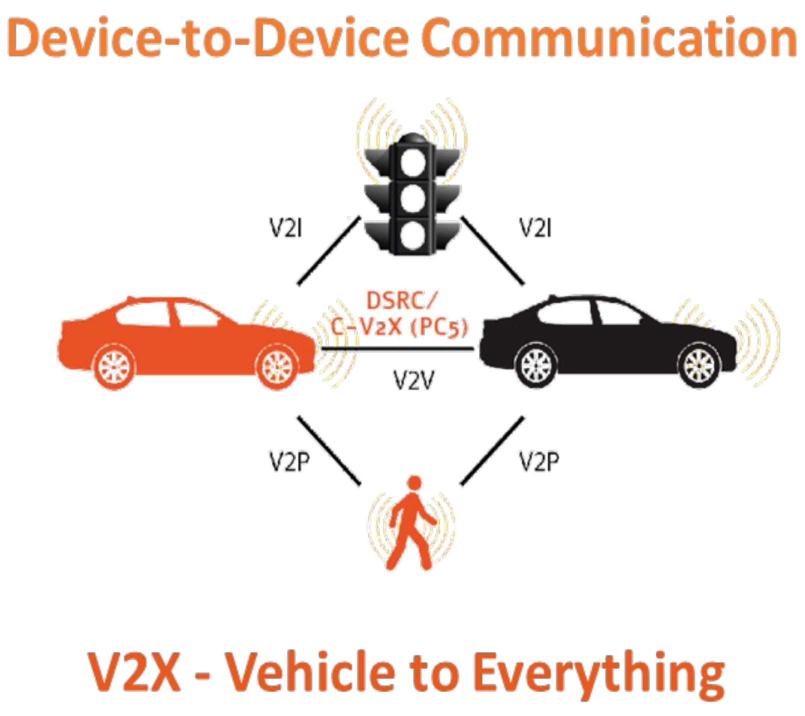


Source: GAO.

- Broadcasts BSMs 10 times per second
- Transmit power are about 100mW (20dBm @Antenna Port - Per IEEE802.11-D.2.2 Transmit power level) with a nominal range of 300m (360^{o} coverage)
- DSRC units share the same channel

C-V2X Basics

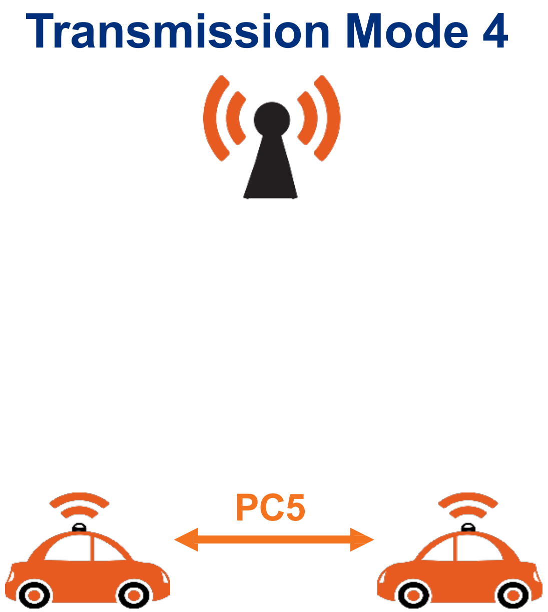
- C-V2X is a V2X radio layer:
  - C-V2X is Device-to-Device (D2D) communication service added to the LTE Public Safety ProSe (Proximity Services) Services
  - C-V2X makes use of the D2D interface – PC5 (aka Side Link) for direct Vehicle-to-Everything communication
  - C-V2X takes the place of DSRC radio layer in relevant regions
  - V2V, V2I and V2P



ITS Layers Remain Unchanged!

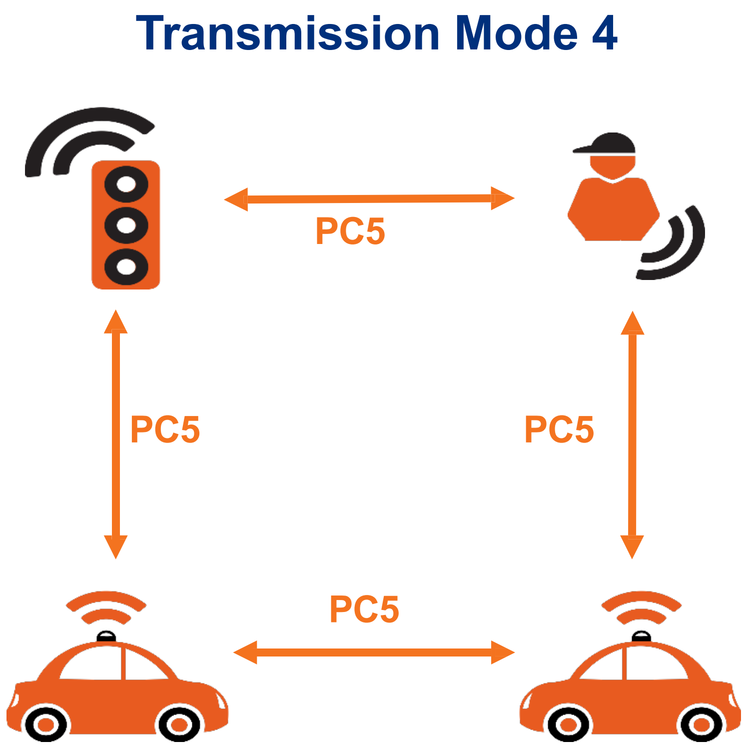
C-V2X Basics

- C-V2X Transmission Mode 4:
  - **Mode 4** – Stand alone, distributed
  - Uses GNSS for location and time for synchronization



C-V2X Basics

- Transmission Mode 4:
  - Out of Coverage operation: The transmitting vehicle is not connected to the network
  - No SIM card or inter-operator collaboration is required
  - Each vehicle performs its own scheduling and allocation
  - No dependency on inter-vehicle components (eNB, Allocation Server etc…)
  - Mandatory for SAE, ETSI



C-V2X Air Interface

- C-V2X is based on LTE (4G) uplink transmission - SC-FDMA (Single Carrier Frequency Division Multiple Access) signal:
  - A single carrier multiple access technique which has similar structure and performance to OFDMA
  - Utilizes single carrier modulation and orthogonal frequency multiplexing using DFT-spreading in the transmitter and frequency domain equalization in the receiver
  - A salient advantage of SC-FDMA over OFDM/OFDMA is low Peak-to- Average Power Ratio (PAPR). Enables efficient transmitter and improved link budget

In Summary

Both Technologies will do the JOB!

But:

- Industry is waiting for regulatory certainty, Government Mandate is preferred!
- C-V2X has to reach automotive production maturity
- Implementation and deployment will depend on OEM system architecture
- The market will demand standalone V2X module for OEMs and aftermarket because V2X is a safety critical sensor.

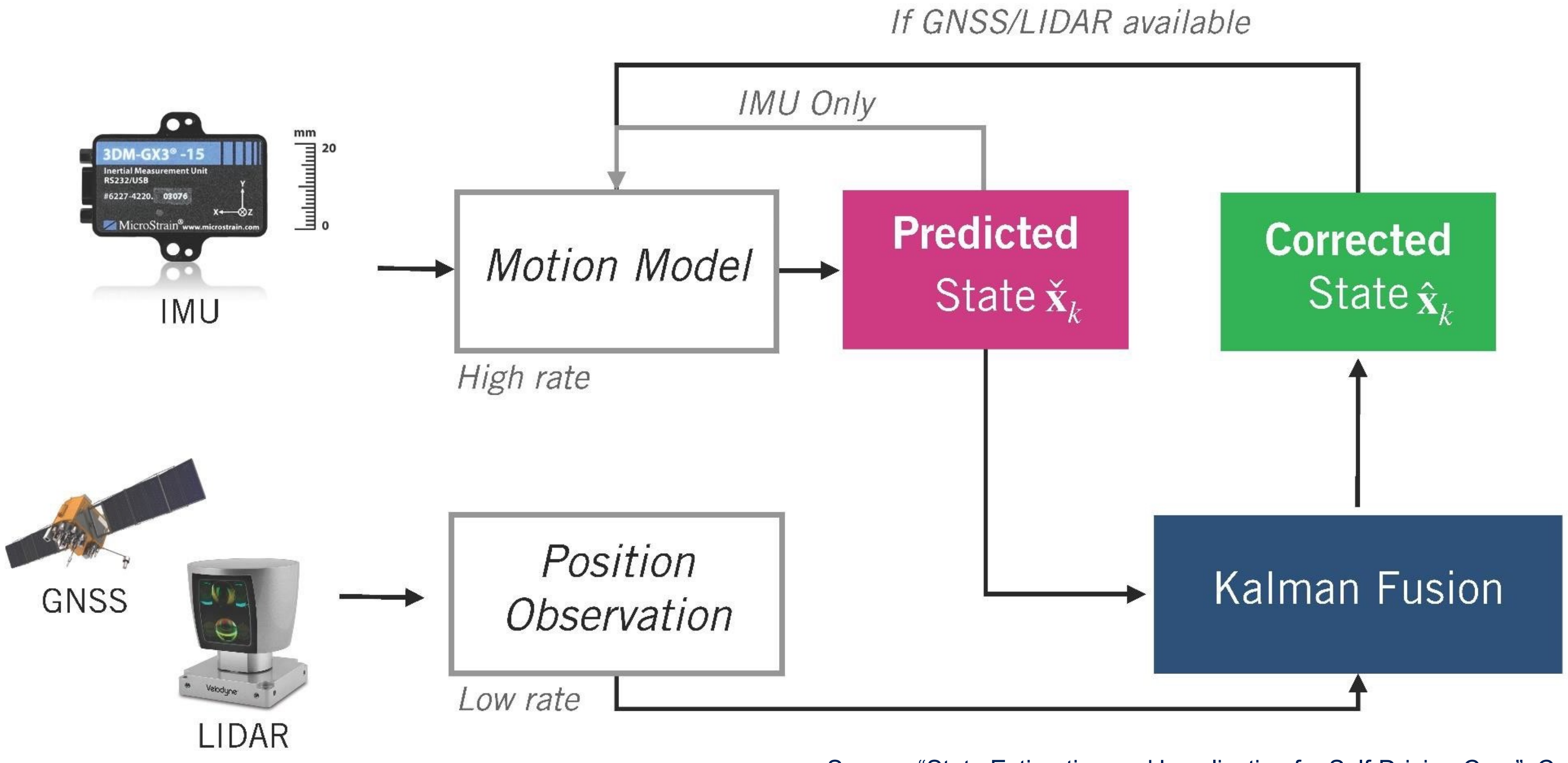
Automotive ADAS Systems

Sensor Fusion Example

Multi-sensor Fusion for State Estimation

Extended Kalman Filter |
IMU + GNSS + LIDAR

This is a rule based fusion example, we will see another fusion later



Source: “State Estimation and Localization for Self-Driving Cars”, Coursera by University of Toronto

PART II: Reducing Human Efforts in Visual Perception

Autonomous Driving Lab, DAMO Academy

CAINIAO 菜鸟 | 達摩院 ALIBABA DAMO ACADEMY

Carrier
Largest Autonomous Driving in logistic

200+ Cities

800+ AutoVehicle

50M+ orders

Truck
Research -> Product

50+ routes across China

30+ test vehicles

100M+km test milage

Heavy Truck
Preliminary Exploration

Built 20+ Auto-Truck

Cainiao, Shentong

Release in 2027

Autonomous Driving Vehicle Is Also A Robot

Autonomous Driving
Understand and Act in 3D World

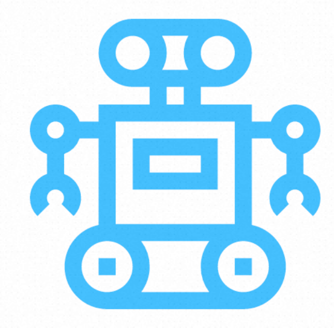
Bus

Taxi

Heavy Truck

Carrier

Common Framework of Robotic System



Robot!

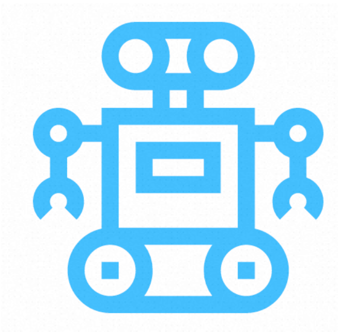
Understand the 3D world

Planning
Data creation

Decide what to do

Control in realistic space
Interact with the world

My Research Focus: Perception + Imagination



Robot

My Research Focus

Perception → Imagination → Decision → Control

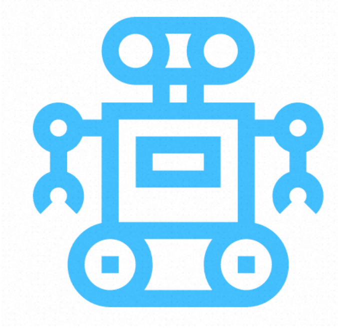
Understand the 3D world

Planning
Data creation

Decide what to do

Control in realistic space
Interact with the world

My Talk Focus: Perception



Robot

My Talk Focus

Perception → Imagination → Decision → Control

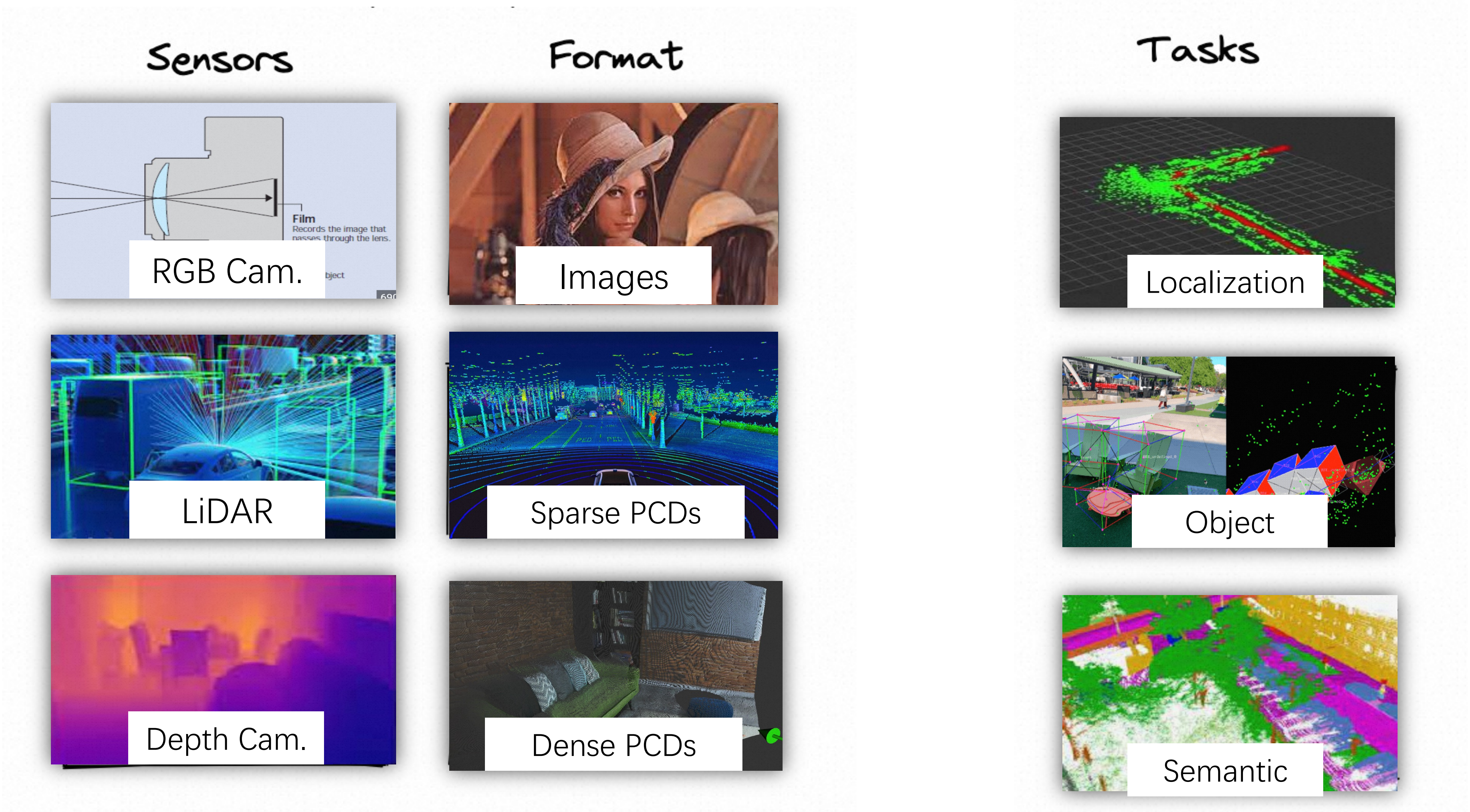
Understand the 3D world

Planning
Data creation

Decide what to do

Control in realistic space
Interact with the world

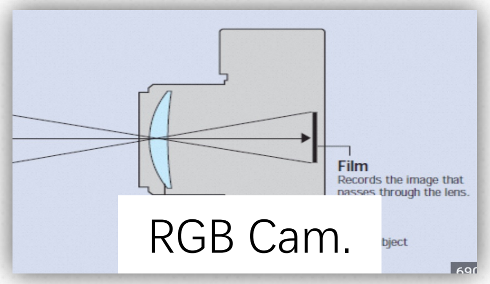
What is Visual Perception?

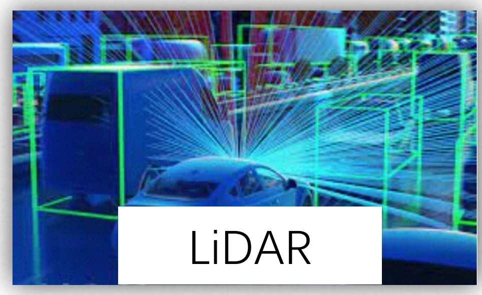


Visual Perception in 3D

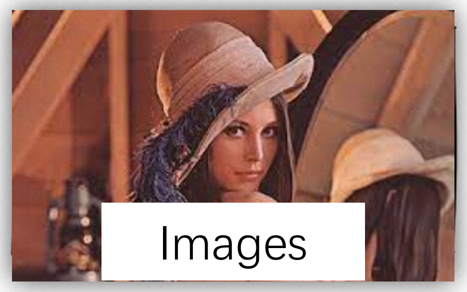
Perception → Imagination → Decision → Control

Sensors



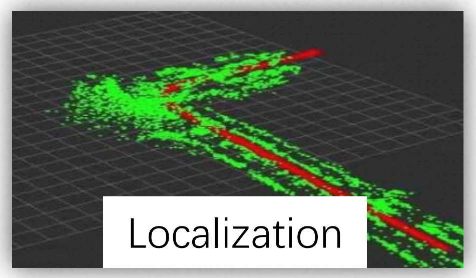


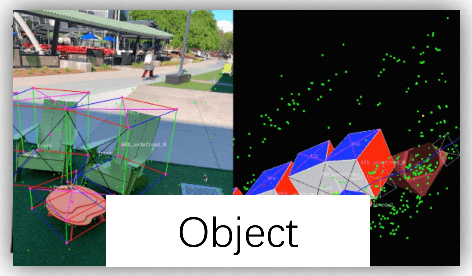
Format

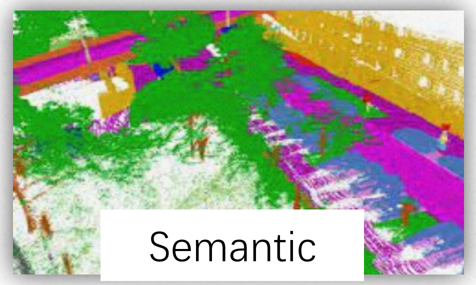


AI Models

Tasks

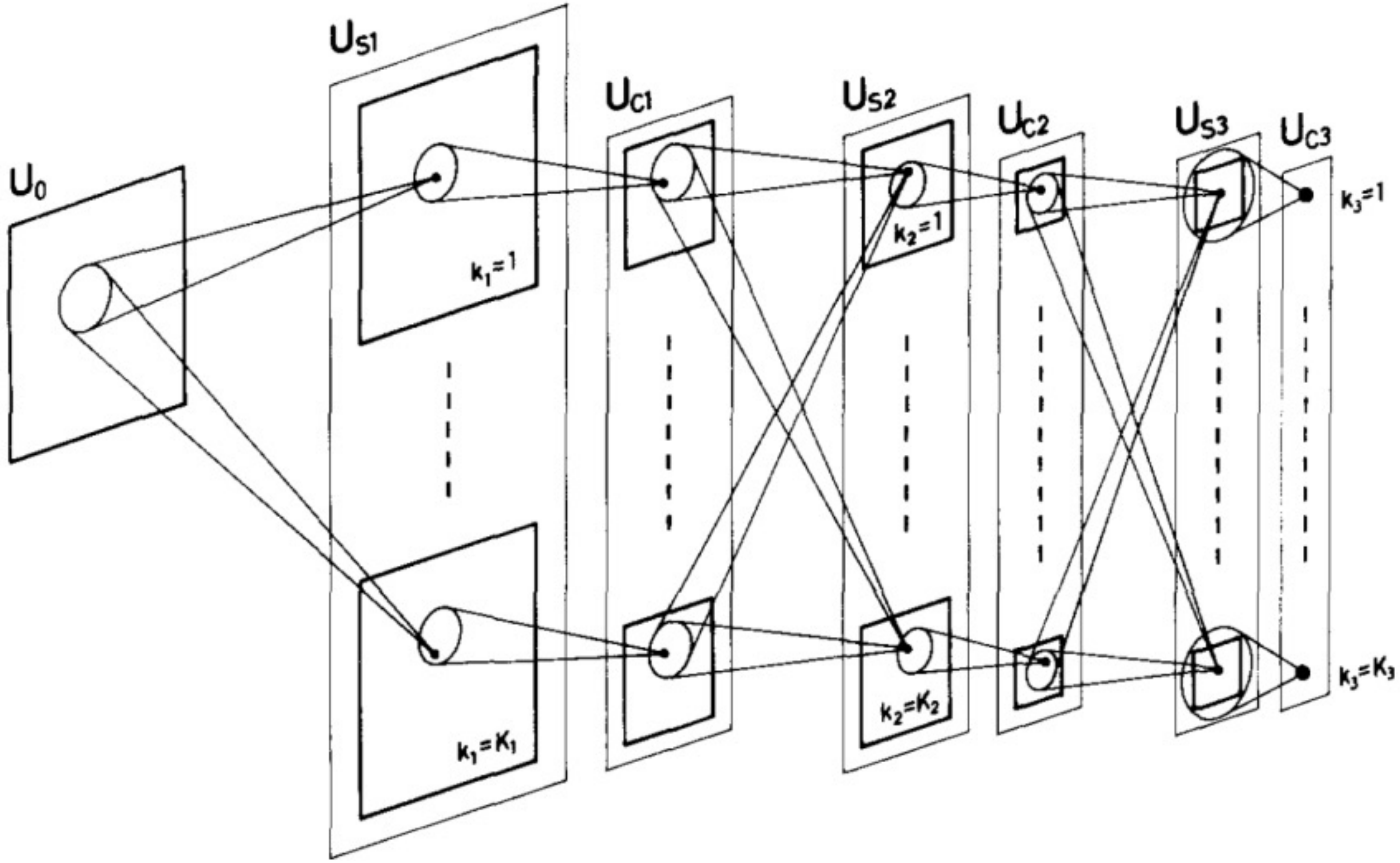






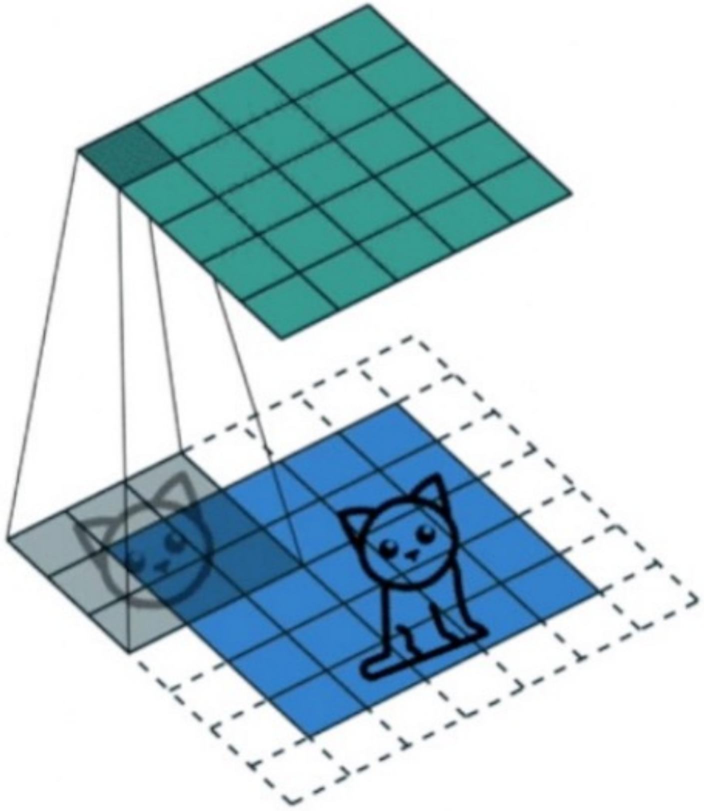
Convolutional neural network

Convolutional Neural Networks

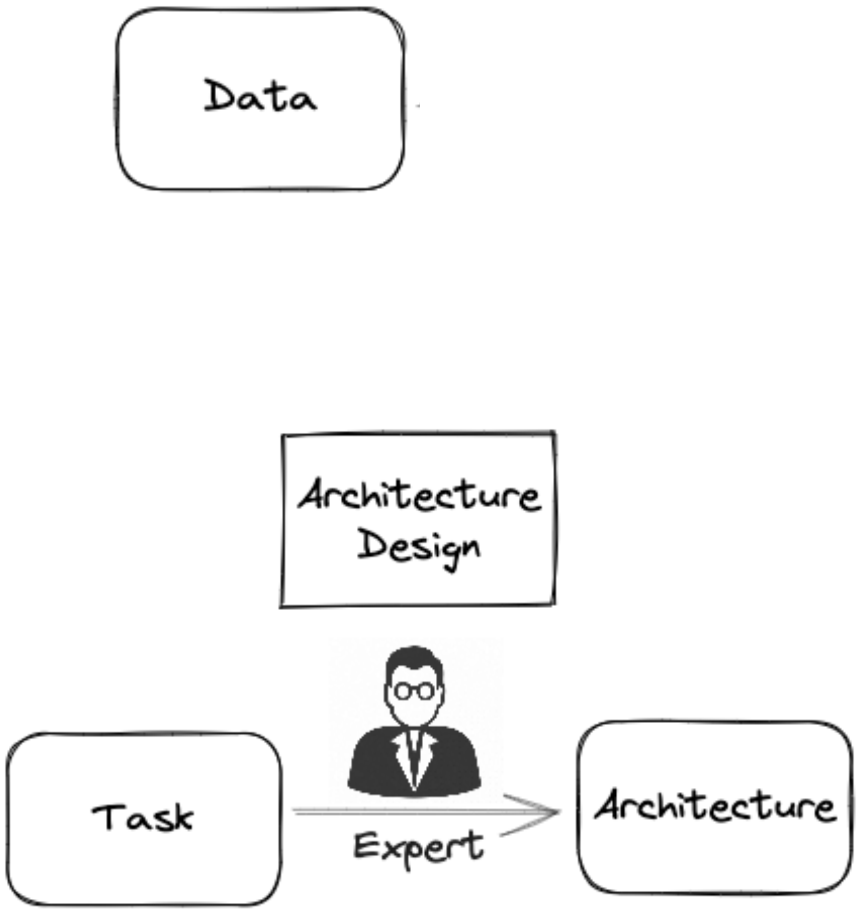


Convolution is template matching …

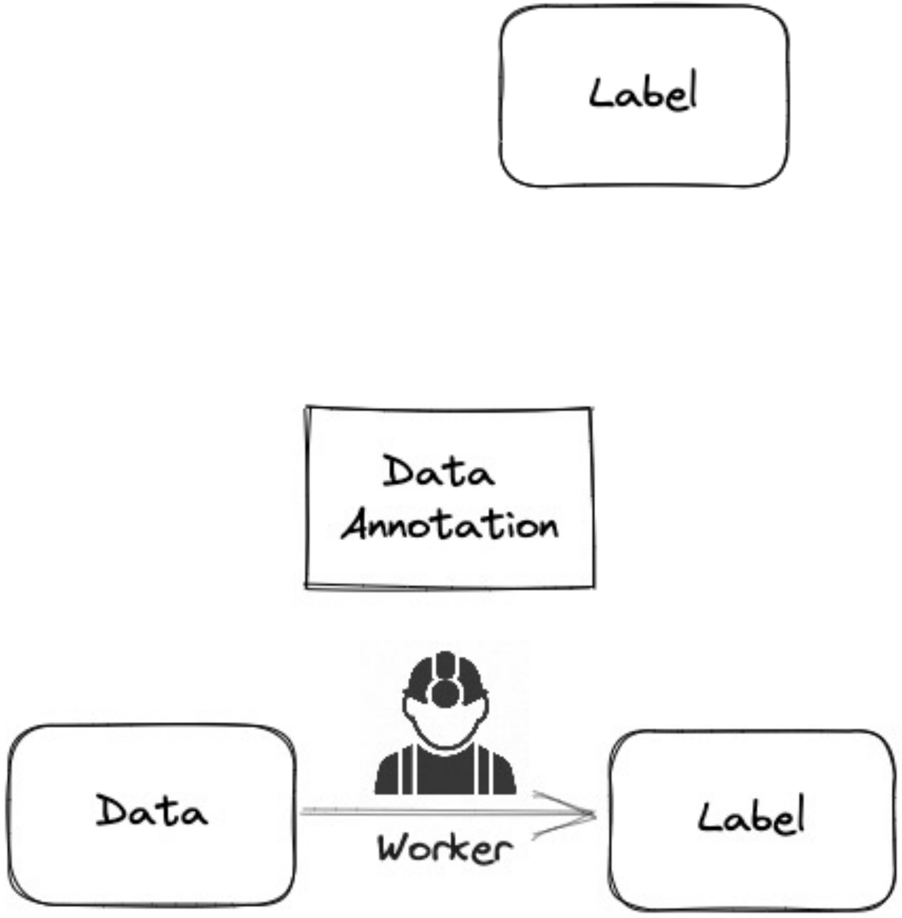
- with a sliding window
- abstract templates
- similarity measured by dot product
- stronger activation, better matching



Supervised Learning in Visual Perception



Manual Design Architectures



Large-scale Annotation

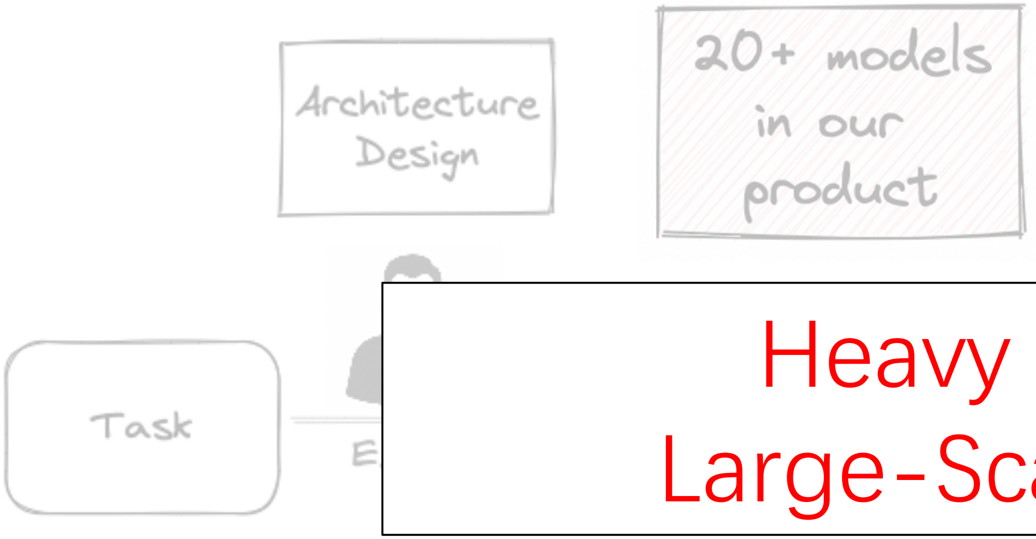
What are Key Challenges in Supervised Visual Perception?

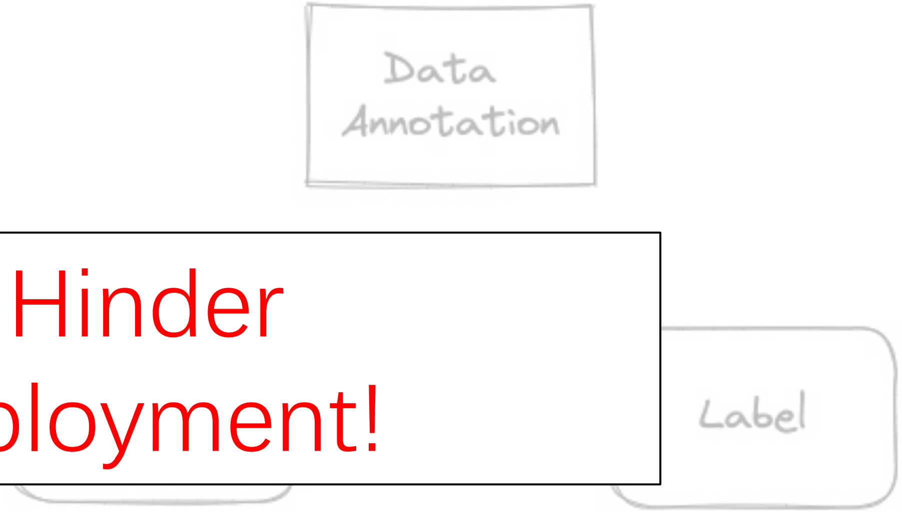
1. Large Efforts in Architecture Design
2. Large Efforts in Data Annotation

Heavy Human Efforts in Visual Perception

Key Challenge 1: Large Efforts in Architecture Design

Key Challenge 2: Large Efforts in Data Annotation





Heavy Efforts Hinder Large-Scale Deployment!

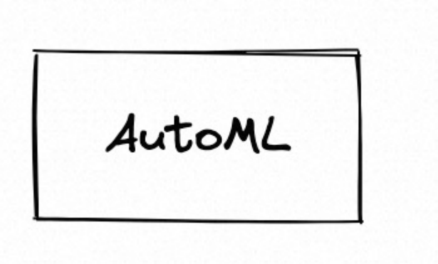
ML Expert
- designing network
- experiments
- maintaining system
- integration and etc.

Cost: 1 Million per person
Output: 1-2 Model per year

3D Data Annotation
- Low unit price
- Large-scale data
- > 10 Million annotation

Company Cost
> 40 Million per year

Reducing Human Efforts in Visual Perception



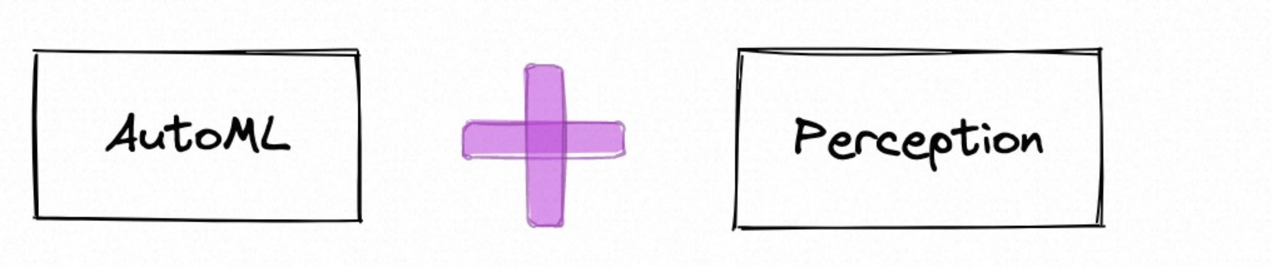
EvalNAS, ICLR 20
LR, CVPR 21
SuperNet, TPAMI 22
...

Address Challenge 1: Large Efforts in Architecture Design

- Identifying why NAS cannot surpass random search
- Our Landmark Regularization solution to address

We will not cover it in this lecture

Reducing Human Efforts in Visual Perception



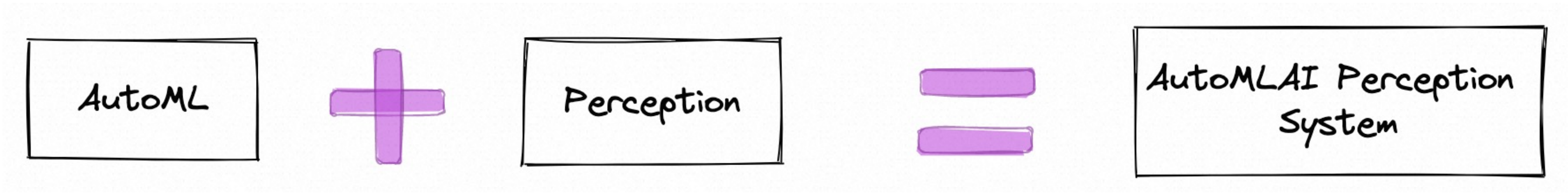
EvalNAS, ICLR 20
LR, CVPR 21
SuperNet, TPAMI 22
...

BEVFusion, NeurIPS 22
BEVHeight, CVPR 23
Rec.UNet, ICCV 19
SMSOP, ECCV 18
...

Address Key Challenge 2: Large Efforts in Data Annotation

- Auto-Labeling and pseudo labels to save human efforts
- High-performance and robust 3D perception framework

Reducing Human Efforts in Visual Perception



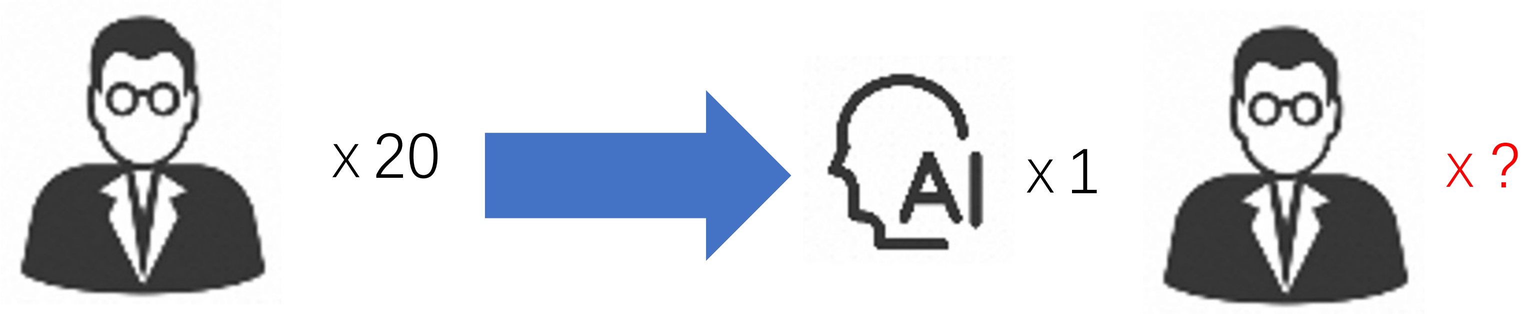
EvalNAS, ICLR 20
LR, CVPR 21
SuperNet, TPAMI 22
...

BEVFusion, NeurIPS 22
BEVHeight, CVPR 23
SMSOP, ECCV 18
...

AI System
- Role: Chief Architect
- Broader AutoML
- Deployed in Alibaba

Address Key Challenges 1 & 2:

- Address both challenges together
- A platform to integrate our latest research advances



Before

AutoML System V1

Key Challenge 1: Large Efforts in Architecture Design

Key Challenge 2: Large Efforts in Data Annotation

Perception in 3D World

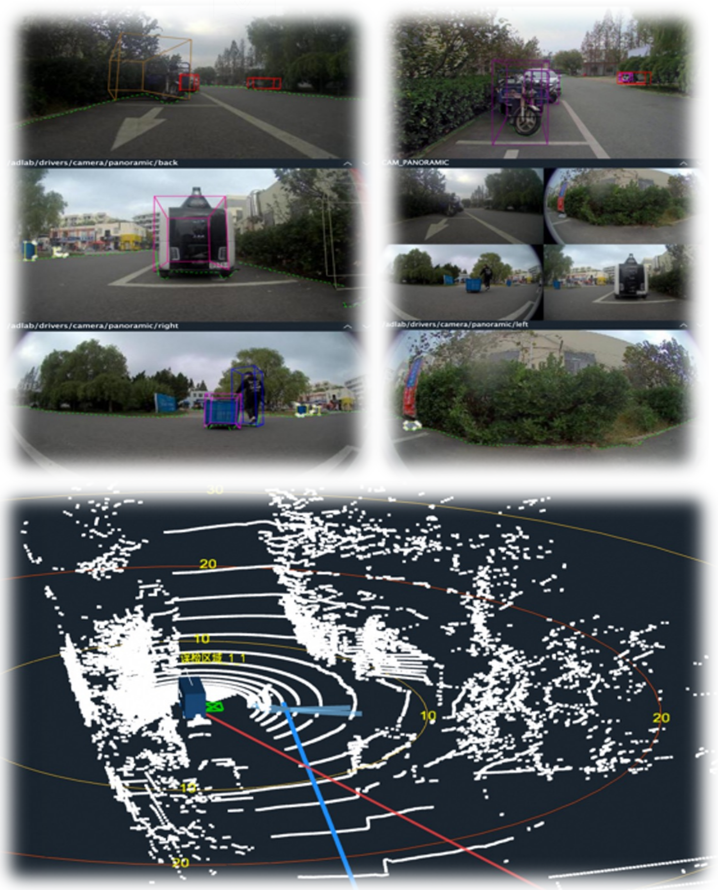
AutoML + Perception = AutoMLAI Perception System

Here

Perception in 3D Understanding

Sensor Data

Camera LiDAR Radar etc.

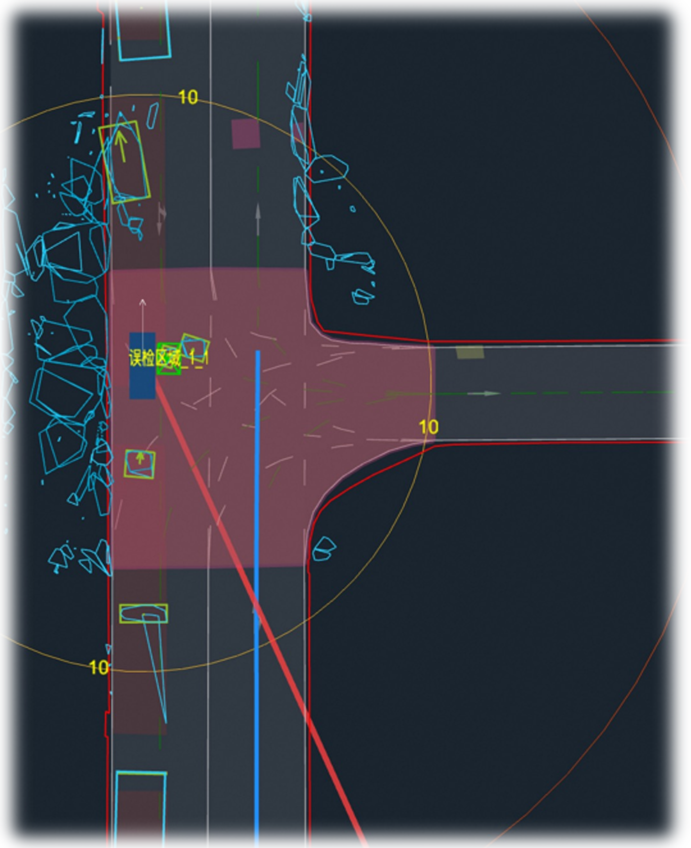


Perception

- Brain of robotics
  - Similar to human
- The only approach to understand the world!
  - Data centric
  - Deep Neural Networks

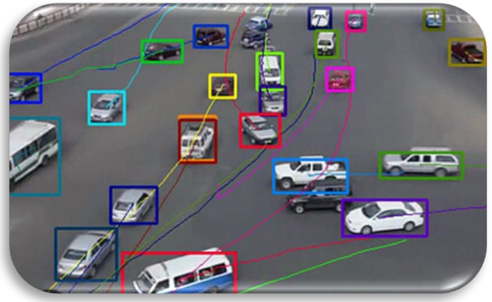
Vectorized space

3D digital world

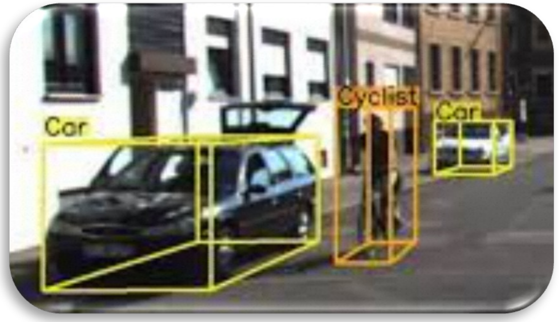


3D Understanding Tasks

Perception

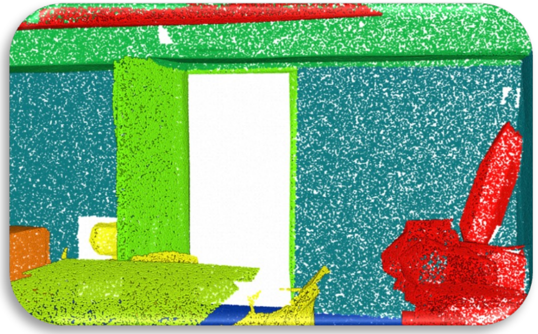


Multi-object Tracking

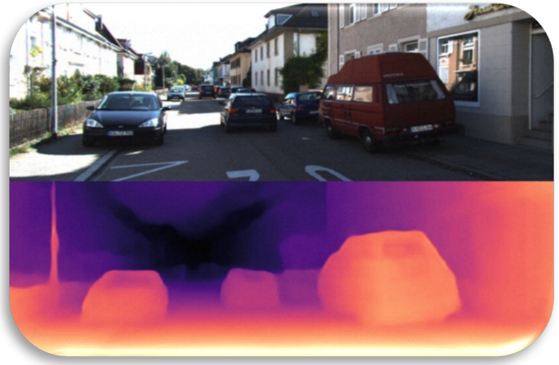


Object Detection

...



Point-cloud Segmentation



Depth Completion

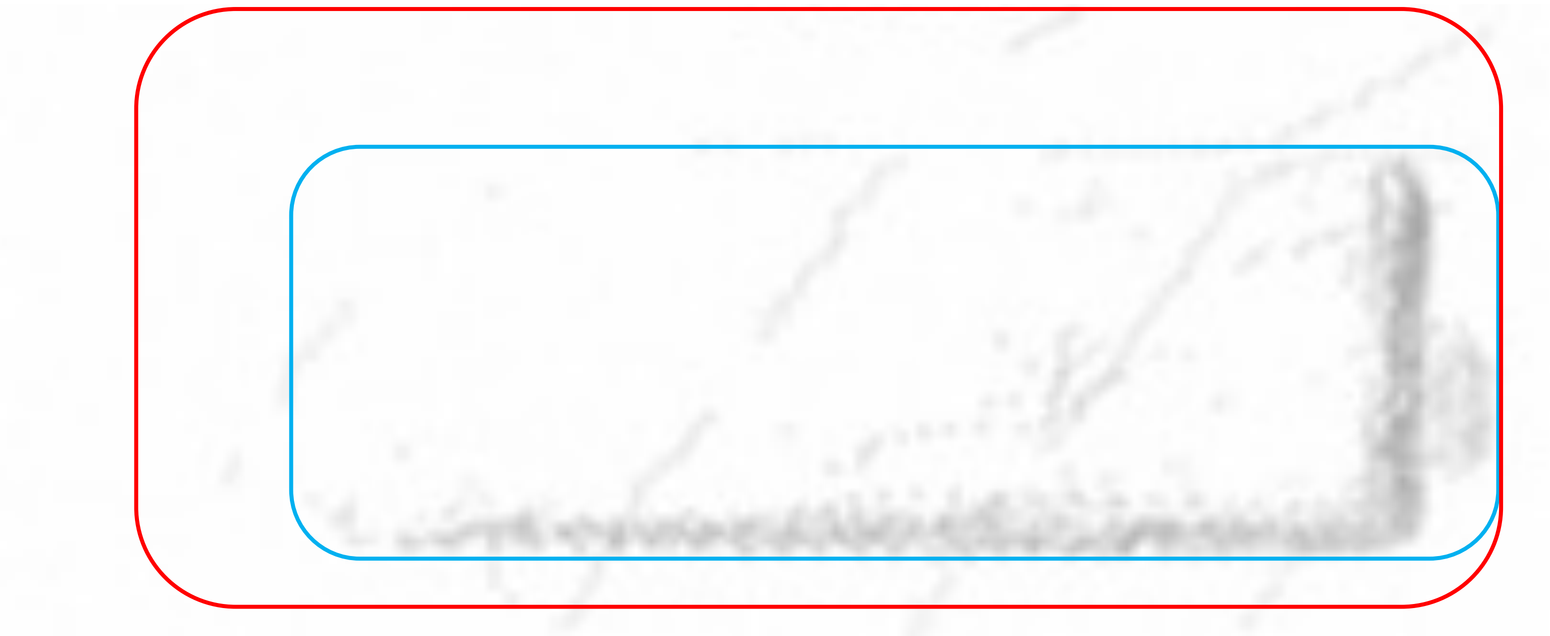
Why 3D Annotation with Multi-sensor Data Is Hard?

Red: GroundTruth

Example of 2D Object Box Annotation

Why 3D Annotation With Multi-sensor Data Is Hard?

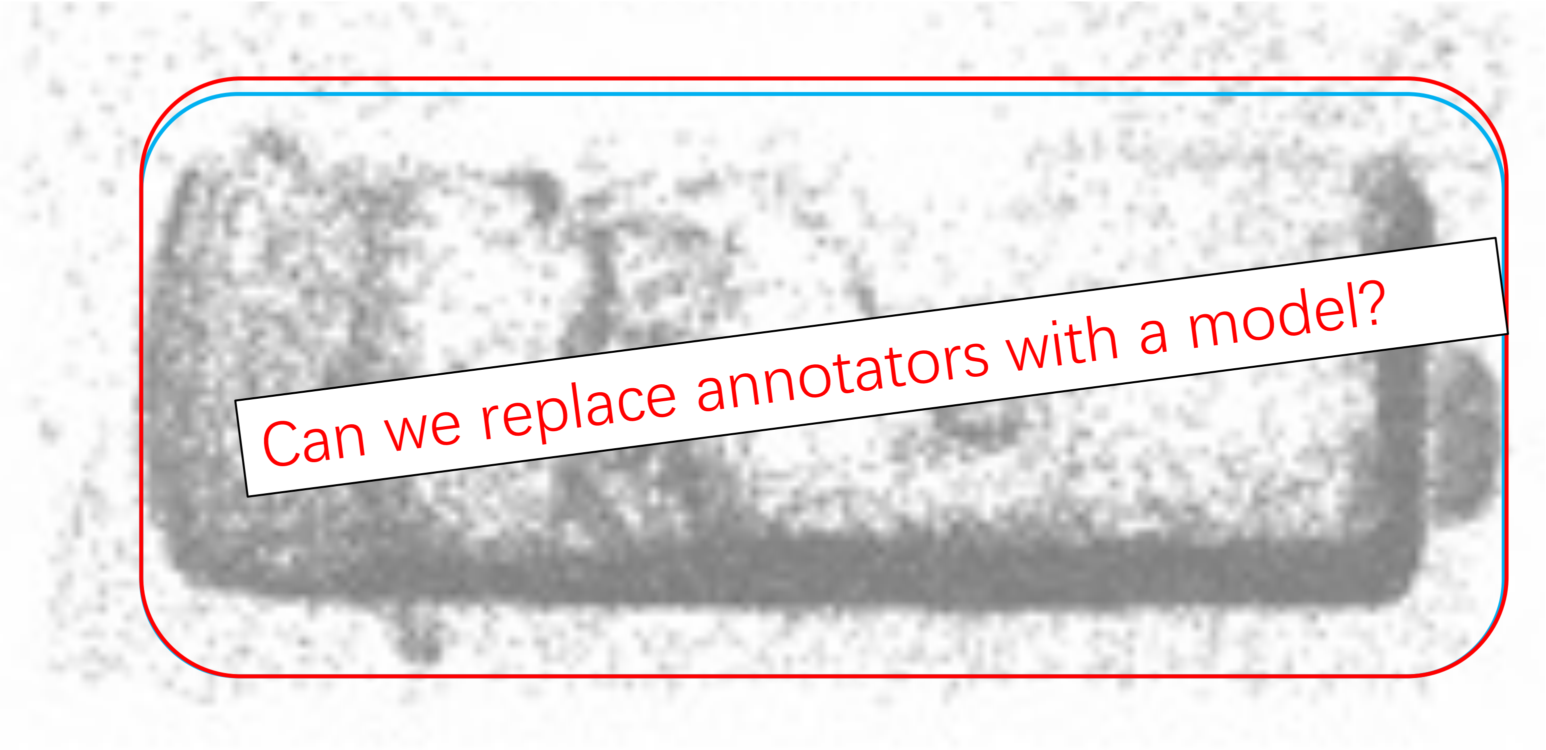
Red: GroundTruth
Blue: Common annotator



Example of 3D Object Box Annotation
(Bird eye view of 3D point clouds)

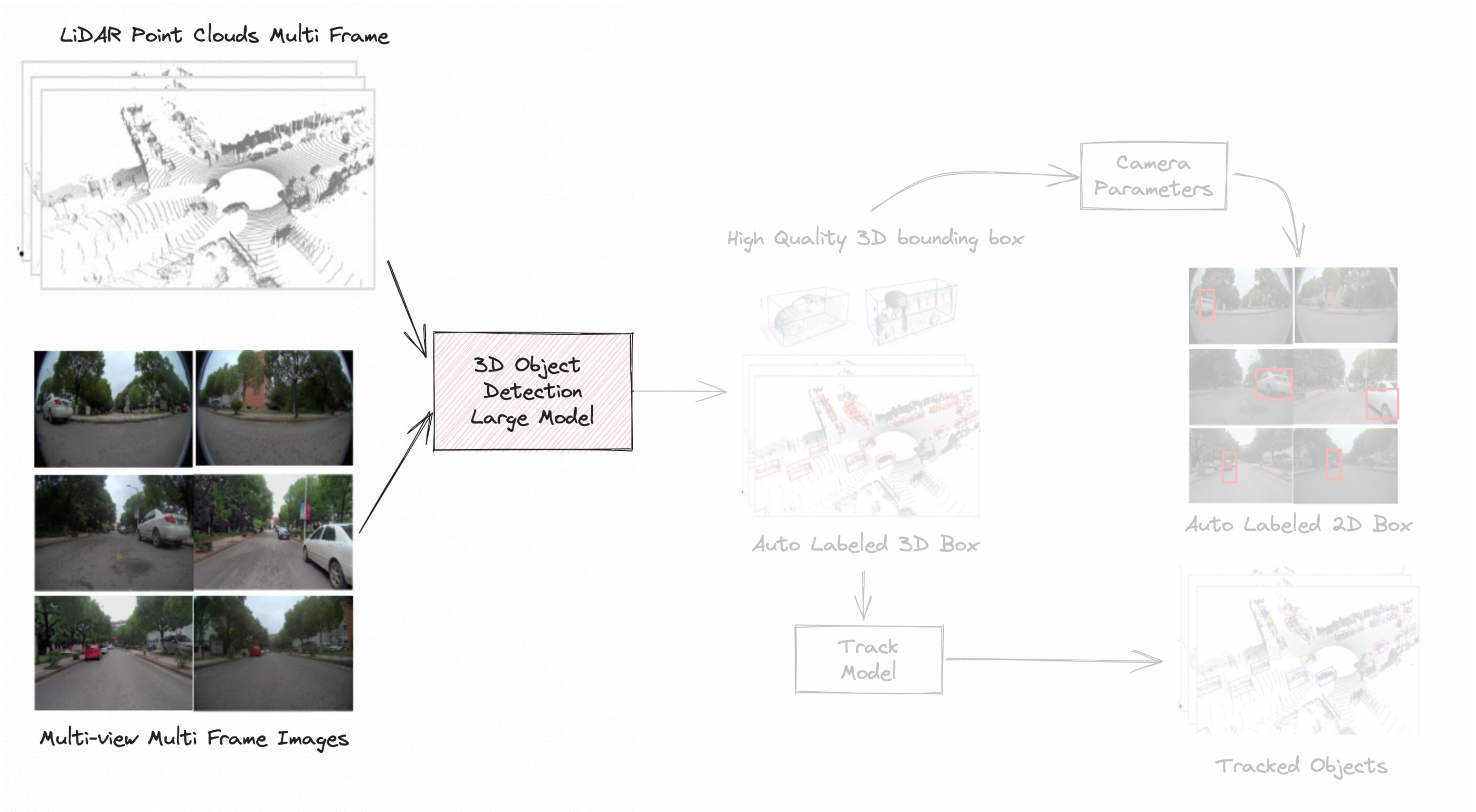
Why 3D Annotation With Multi-sensor Data Is Hard?

Red: GroundTruth
Blue: Common annotator

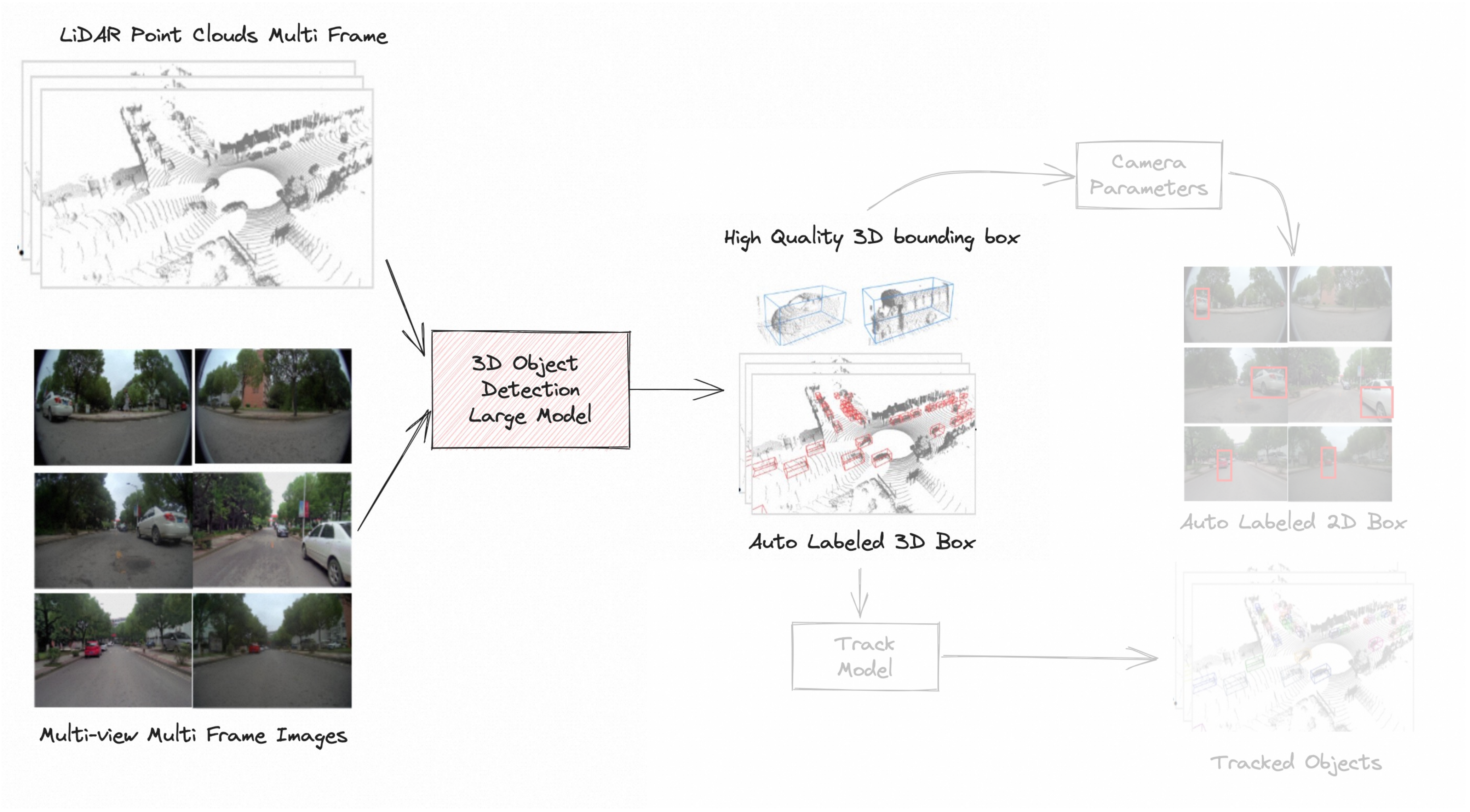


Example of 3D Object Box Annotation
(Bird eye view of 3D point clouds)
Aggregating 100+ frames!

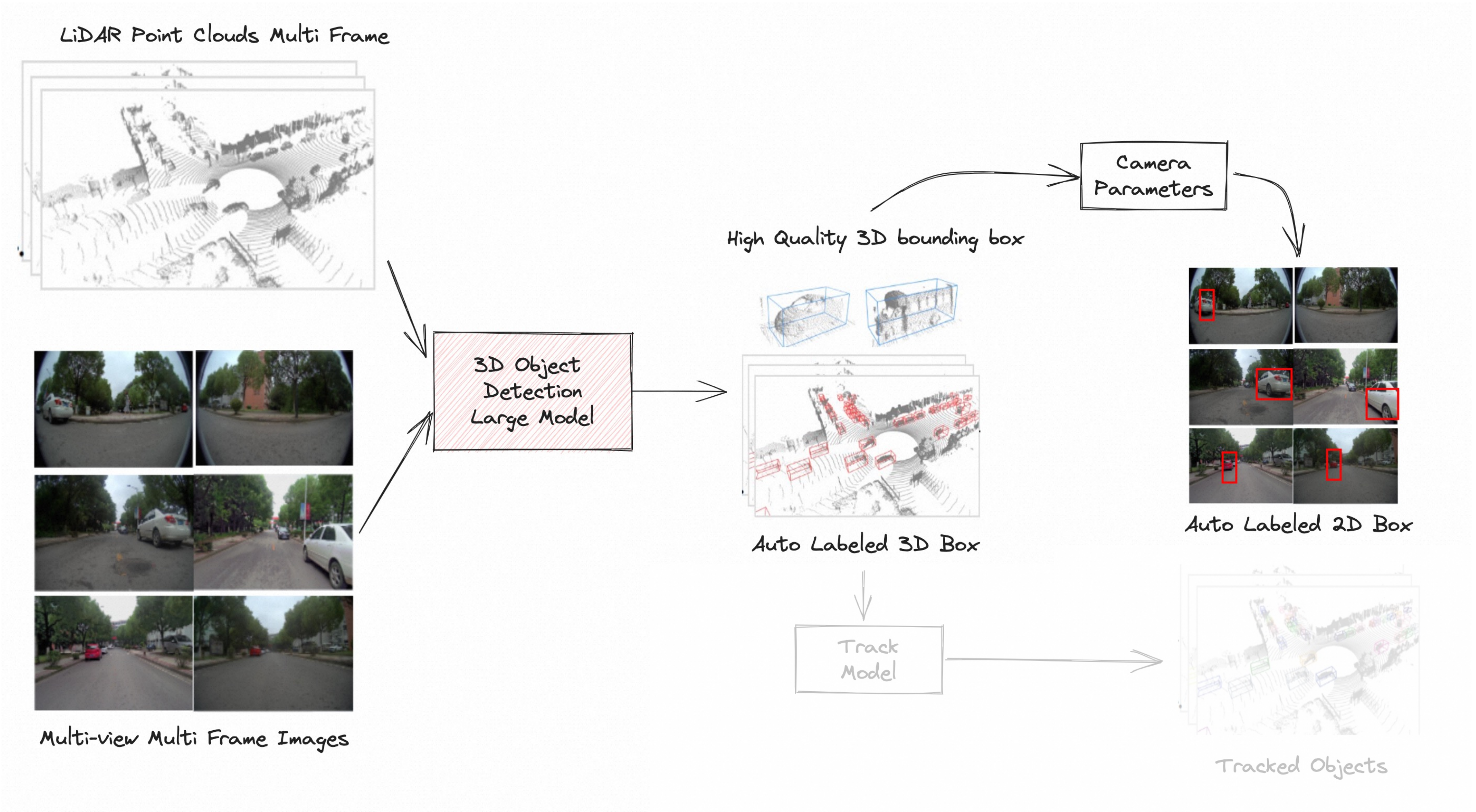
AutoLabel System: Large model as Pseudo Labeler



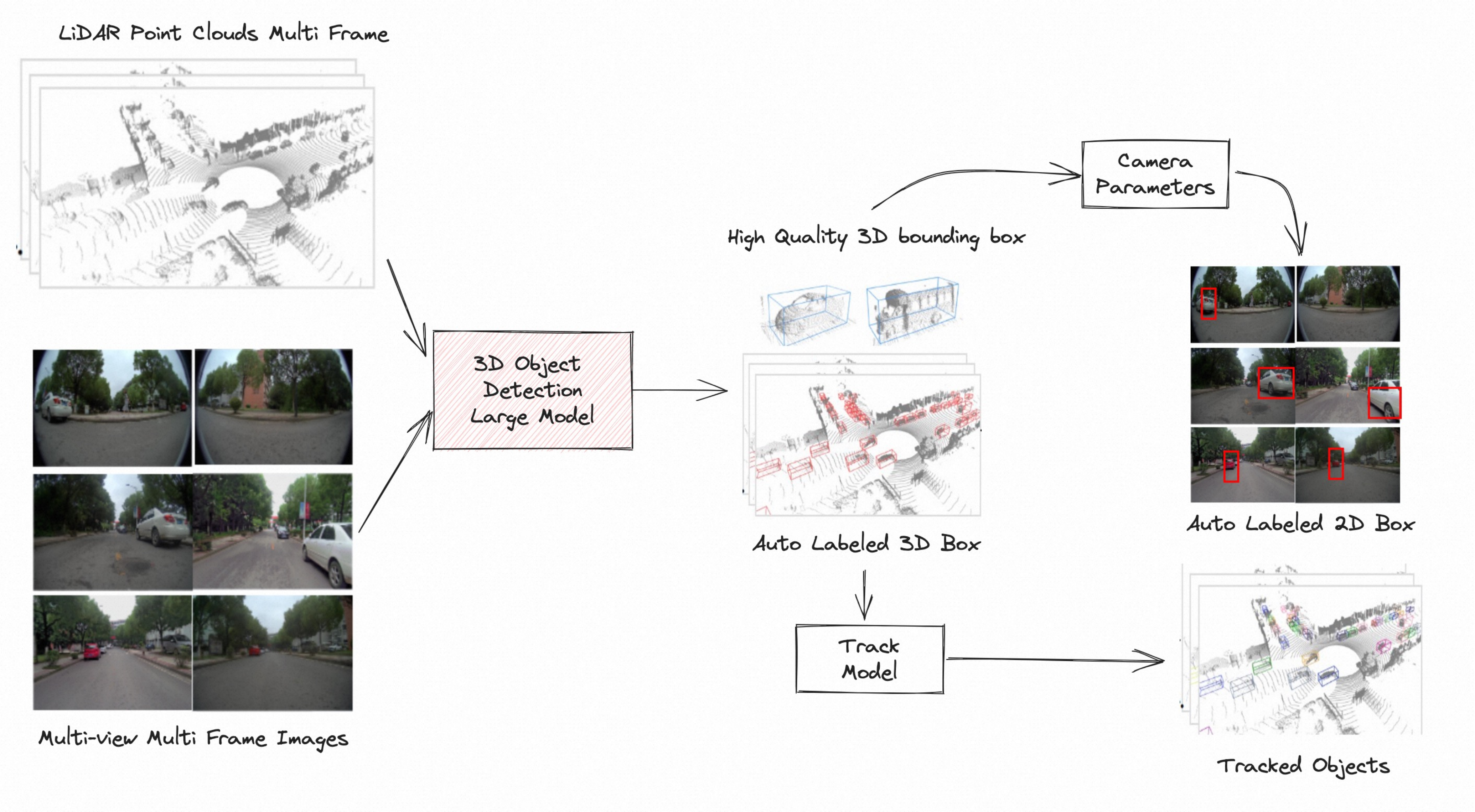
AutoLabel System: Large Model as Pseudo Labeler



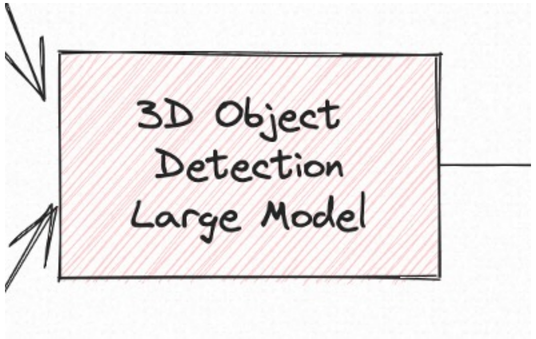
AutoLabel System: Large Model as Pseudo Labeler



AutoLabel System: Large Model as Pseudo Labeler

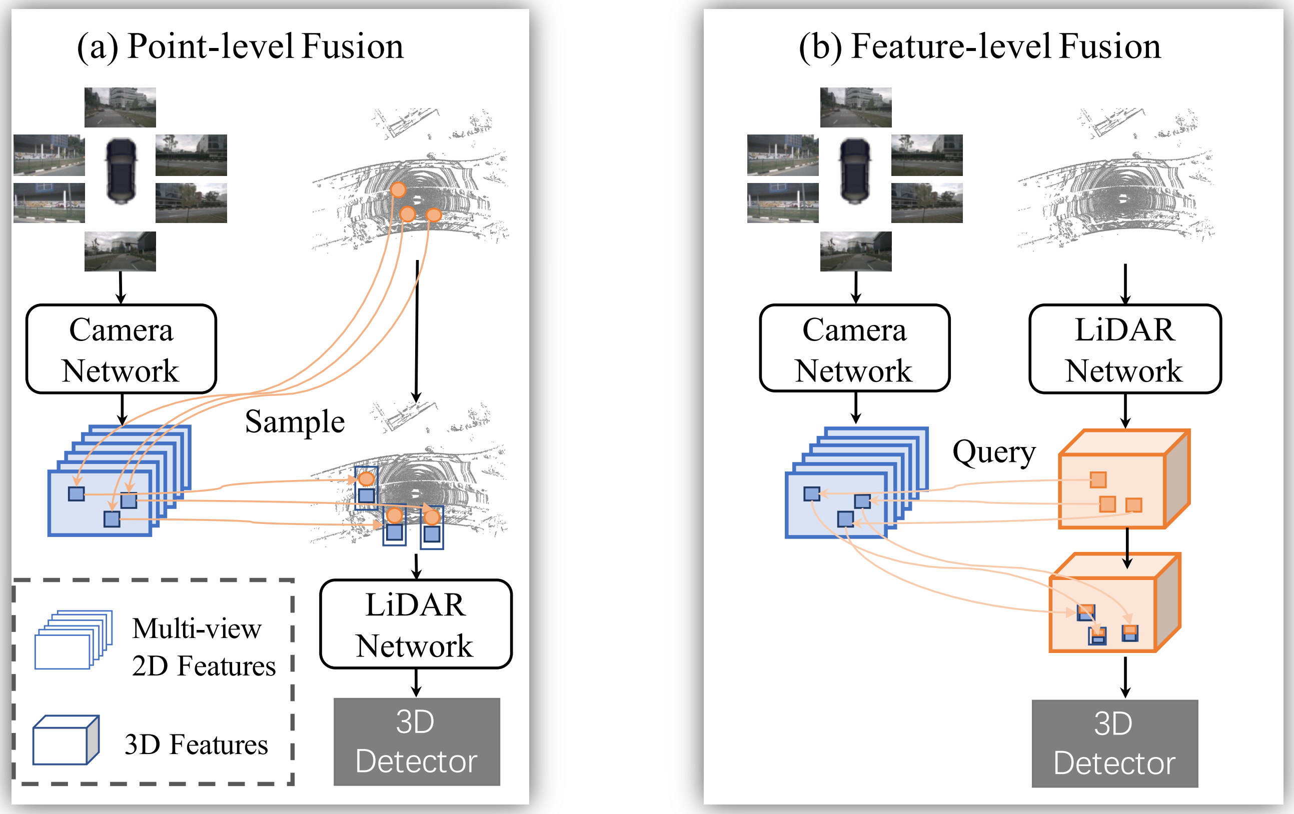


AutoLabel System: Large Model as Pseudo Labeler



Better Base Model	=	Reduce Human Efforts

State of The Art Multi-modality Base Model

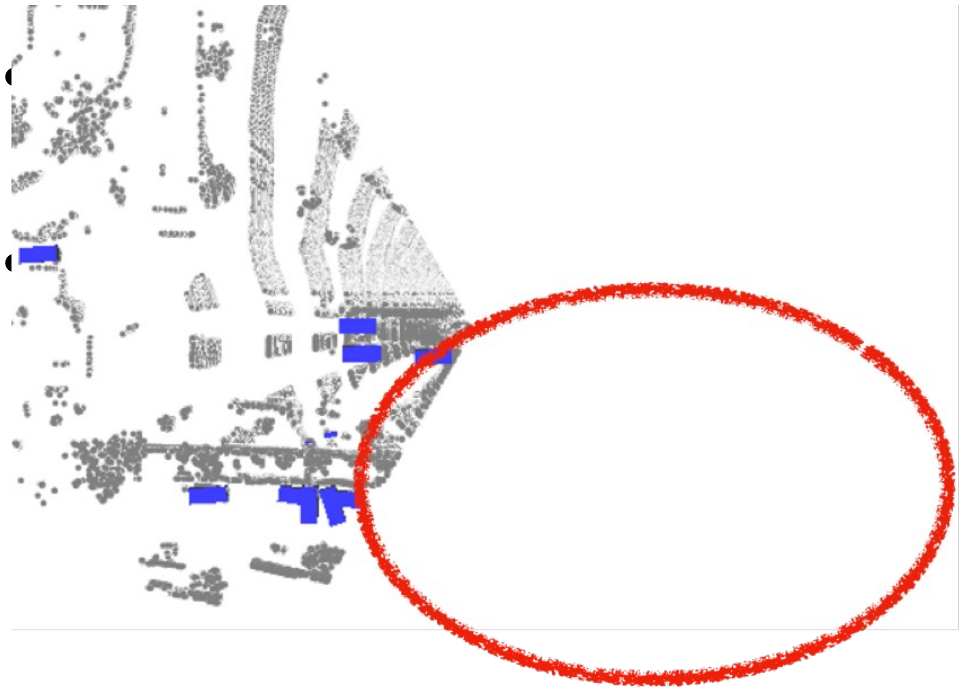


Existing Frameworks of camera-lidar fusion

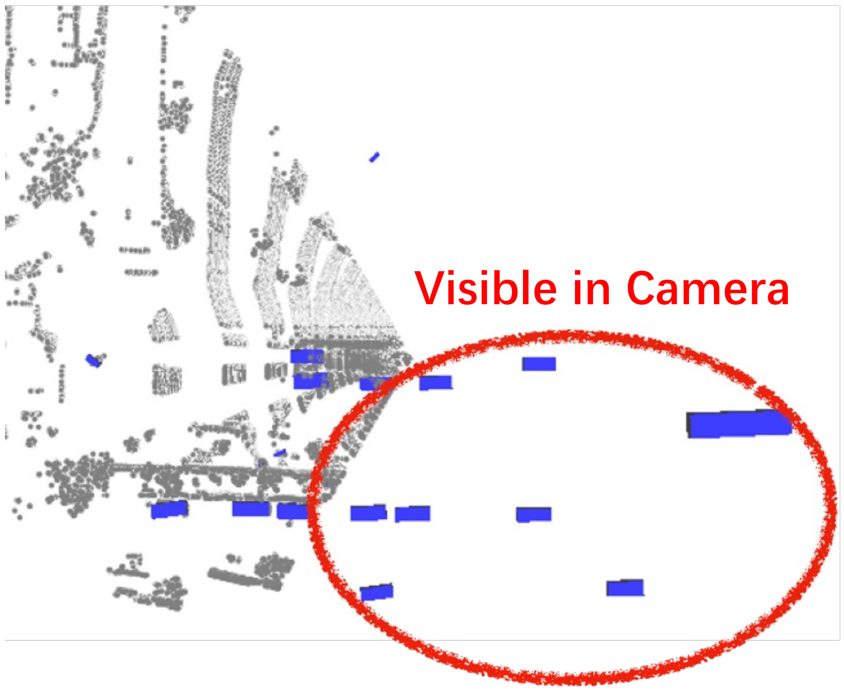
- Fusion starts from point clouds, what if LiDAR fails?

[1] Yu et al. Robustness benchmark of camera-lidar fusion in autonomous driving. CVPR'23 Dataset Paper

SoTA Base Model Fails w/o LiDAR Input



Predictions

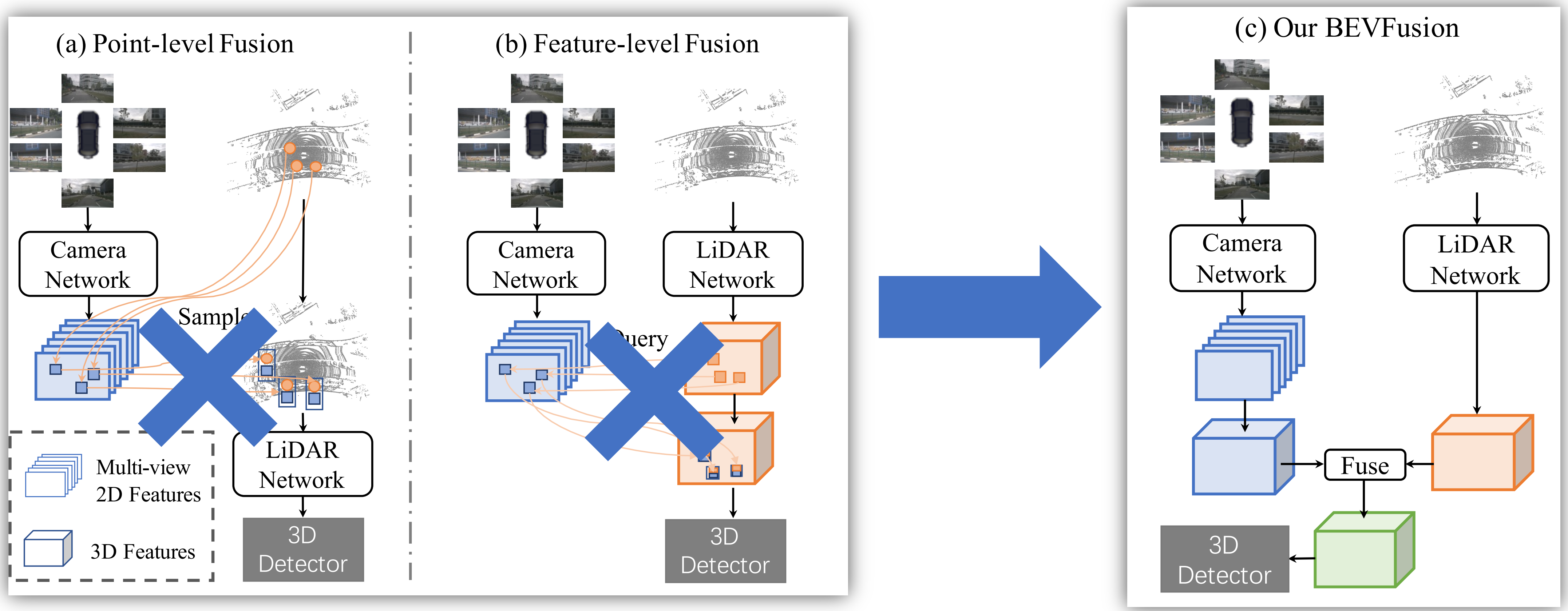


Ground-truth

- Base model with 2 modalities **should not fail** when 1 missing

[1] Yu et al. Robustness benchmark of camera-lidar fusion in autonomous driving. CVPR'23 Dataset Paper

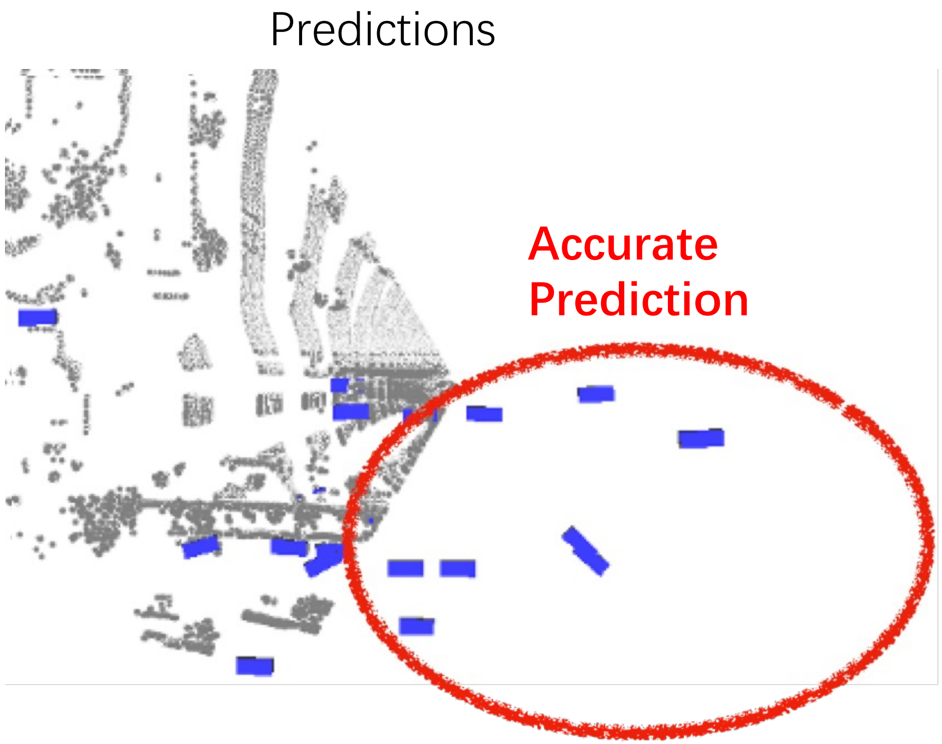
BEVFusion: A Simple yet Robust Base Model Framework

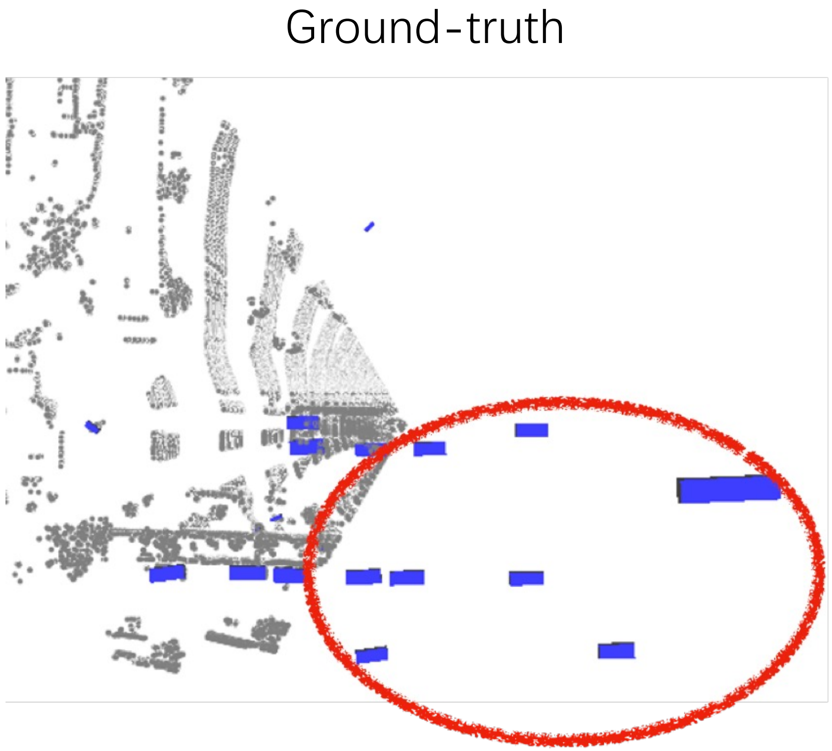


Existing Frameworks of camera-lidar fusion

[1] Liang et al. BEVFusion: A simple yet robust framework for camera-lidar fusion in 3D detection. NeurIPS'22, Spotlight, Supervised intern.

Our BEVFusion Framework is Robust to LiDAR Failure





- The first robust framework that is agnostic to LiDAR failure
- **+30 mAP** compared to baselines
- Become a de-facto standard
- Many follow ups (MetaBEV, BEVFusion 4D, etc.)

[1] Liang et al. BEVFusion: A simple yet robust framework for camera-lidar fusion in 3D detection. NeurIPS'22, Spotlight, Supervised intern.

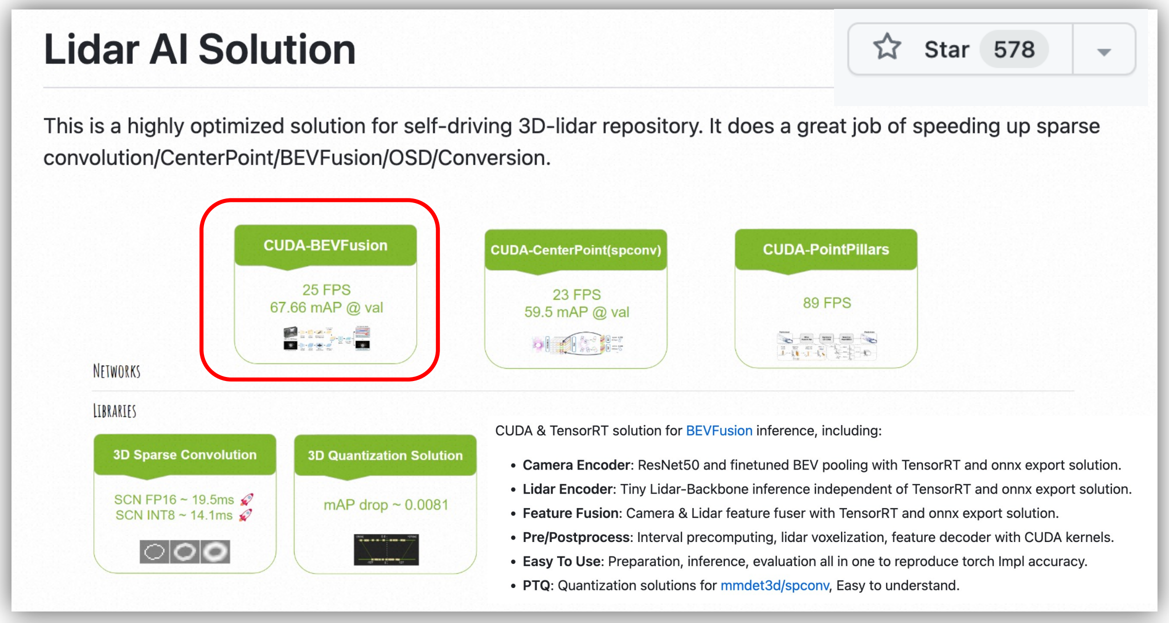
BEVFusion Deployed in Alibaba

High-Quality Ground-truth	Labeler Army	v.s.	Auto Label	
Accuracy (mIoU)	83.12		91.35	(8.23+)
Time (per box)	25s		0.005s	(5000x faster)
Cost (per box)	1 RMB		0.0001 RMB	(10000x cheaper)

- BEVFusion + AutoLabel system surpasses human level annotation!
  - By a large margin

[1] Liang et al. BEVFusion: A simple yet robust framework for camera-lidar fusion in 3D detection. NeurIPS'22, Spotlight, Supervised intern.

BEVFusion Other Impact



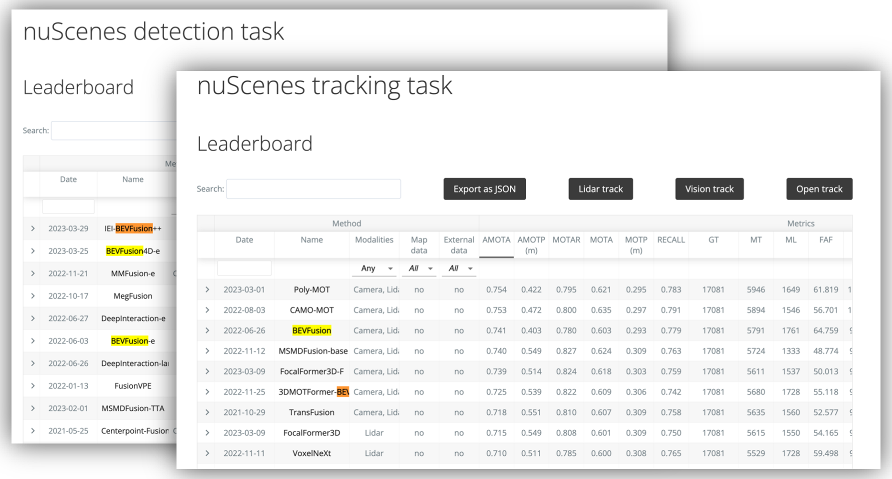
Lidar AI Solution

Star 578

This is a highly optimized solution for self-driving 3D-lidar repository. It does a great job of speeding up sparse convolution/CenterPoint/BEVFusion/OSD/Conversion.

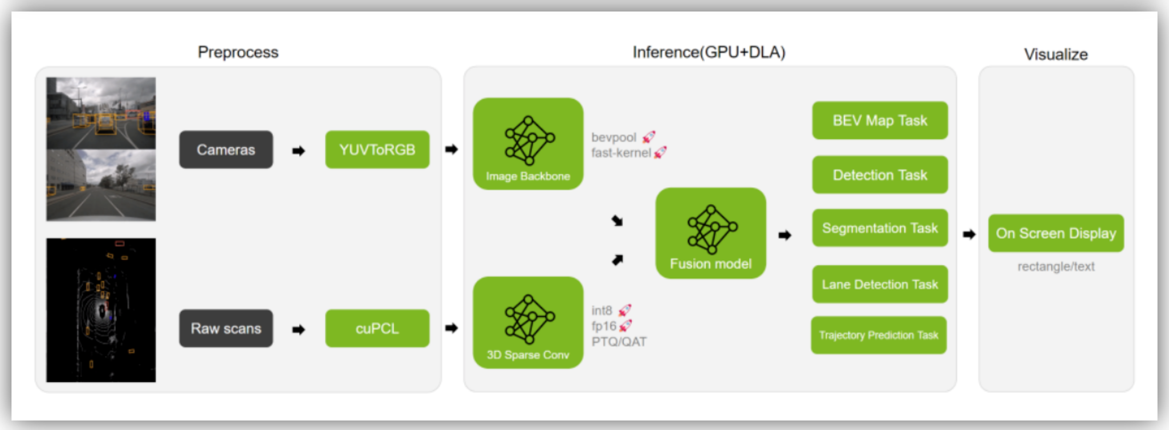
CUDA & TensorRT solution for BEVFusion inference, including:

- **Camera Encoder**: ResNet50 and finetuned BEV pooling with TensorRT and onnx export solution.
- **Lidar Encoder**: Tiny Lidar-Backbone inference independent of TensorRT and onnx export solution.
- **Feature Fusion**: Camera & Lidar feature fuser with TensorRT and onnx export solution.
- **Pre/Postprocess**: Interval precomputing, lidar voxelization, feature decoder with CUDA kernels.
- **Easy To Use**: Preparation, inference, evaluation all in one to reproduce torch Impl accuracy.
- **PTQ**: Quantization solutions for mmdet3d/spconv, Easy to understand.



Date	Name	Modalities	Map data	External data	AMOTA	AMOTP (m)	MOTAR	MOTA	MOTP (m)	RECALL	GT	MT	ML	FAF
2023-03-01	Poly-MOT	Camera, Lid	no	no	0.754	0.422	0.795	0.621	0.295	0.783	17081	5946	1649	61.819
2022-08-03	CAMO-MOT	Camera, Lid	no	no	0.753	0.472	0.800	0.635	0.297	0.791	17081	5894	1546	56.701
2022-06-26	BEVFusion	Camera, Lid	no	no	0.741	0.403	0.780	0.603	0.293	0.779	17081	5791	1761	64.759
2022-11-12	MSMDFusion-base	Camera, Lid	no	no	0.740	0.549	0.827	0.624	0.309	0.763	17081	5724	1333	48.774
2023-03-09	FocalFormer3D-F	Camera, Lid	no	no	0.739	0.514	0.824	0.618	0.303	0.759	17081	5611	1537	50.013
2022-11-25	3DMOTFormer-BE	Camera, Lid	no	no	0.725	0.539	0.822	0.609	0.306	0.742	17081	5680	1728	55.118
2021-10-29	TransFusion	Camera, Lid	no	no	0.718	0.551	0.810	0.607	0.309	0.758	17081	5635	1560	52.577
2023-03-09	FocalFormer3D	Lidar	no	no	0.715	0.549	0.808	0.601	0.309	0.750	17081	5615	1550	54.165
2022-11-11	VoxelNeXt	Lidar	no	no	0.710	0.511	0.785	0.600	0.308	0.765	17081	5529	1728	59.498

Leading in various tracks of leaderboard



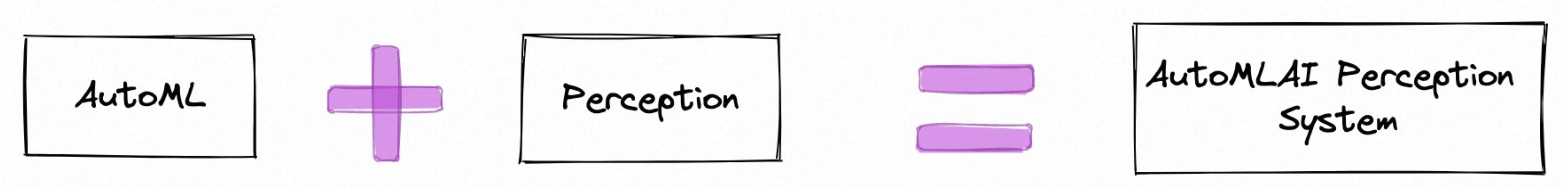
Nvidia Integration as a default AI solution

理想 NIO

HAOMO.AI 達摩院 ALIBABA DAMO ACADEMY HUAWEI

Integration by various AV companies

AI System
ADLab AutoML System



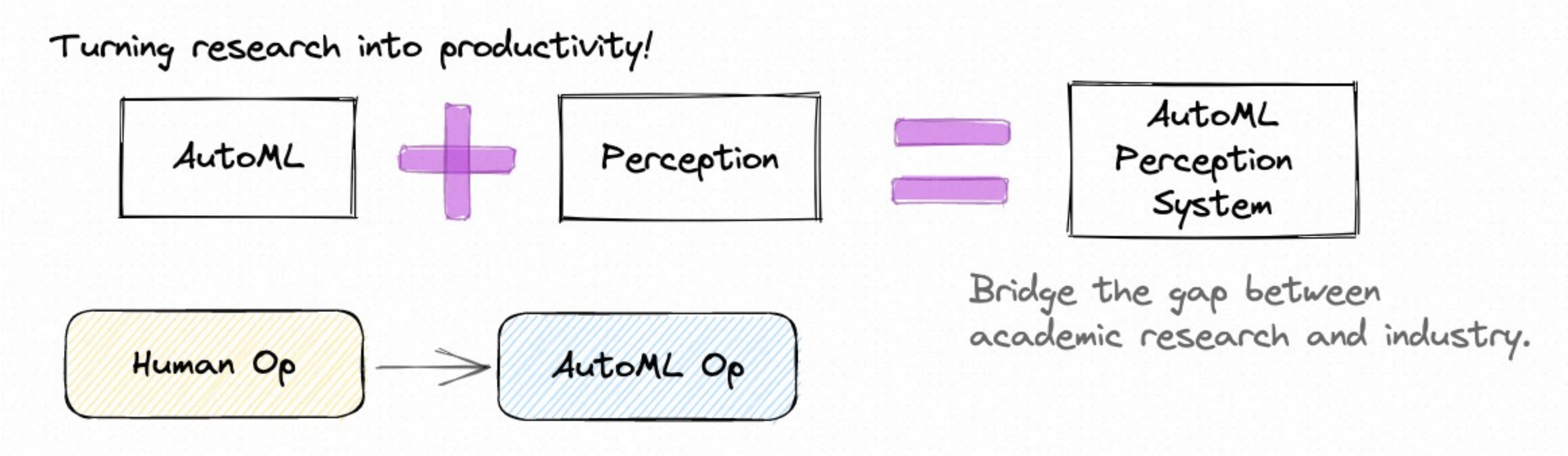
Here

Key Challenge 1: Large Efforts in Architecture Design

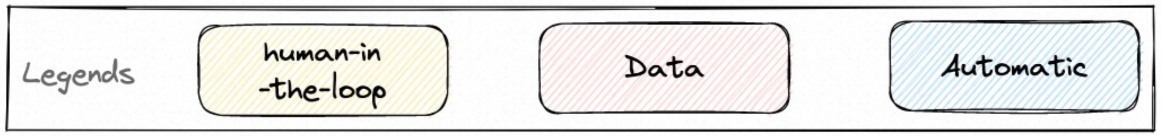
Key Challenge 2: Large Efforts in Data Annotation

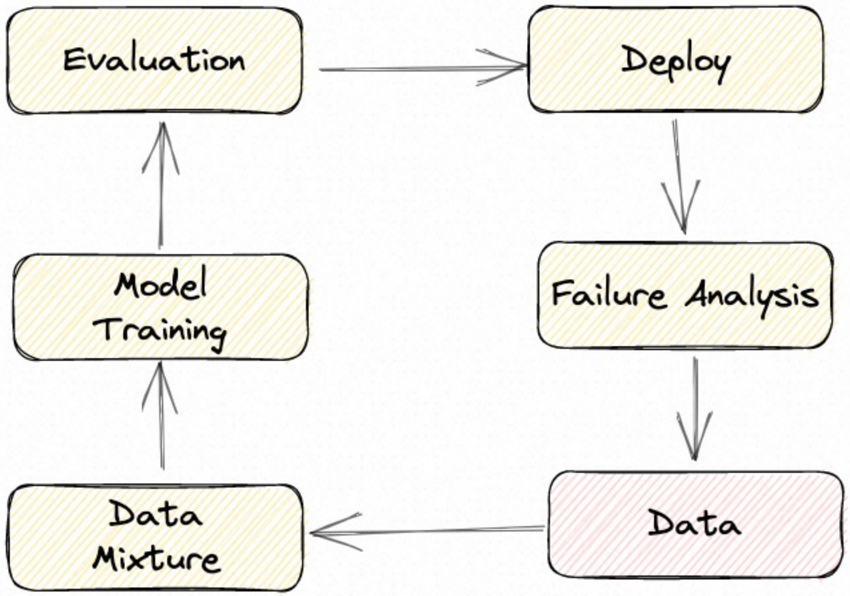
Reducing human efforts by building an AI System

- Automatic machine learning as a system
- My Role: Chief architect



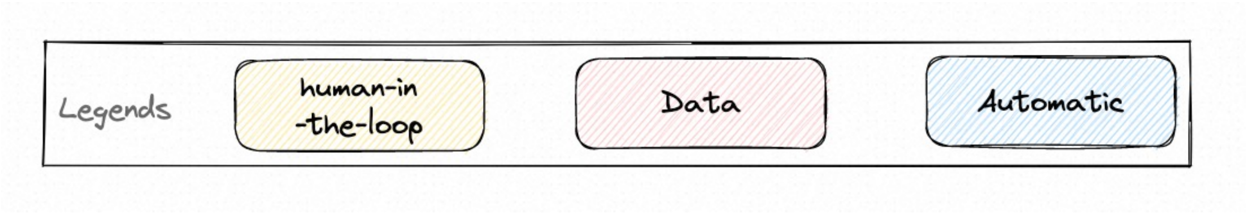
Manual update of an existing deep learning model

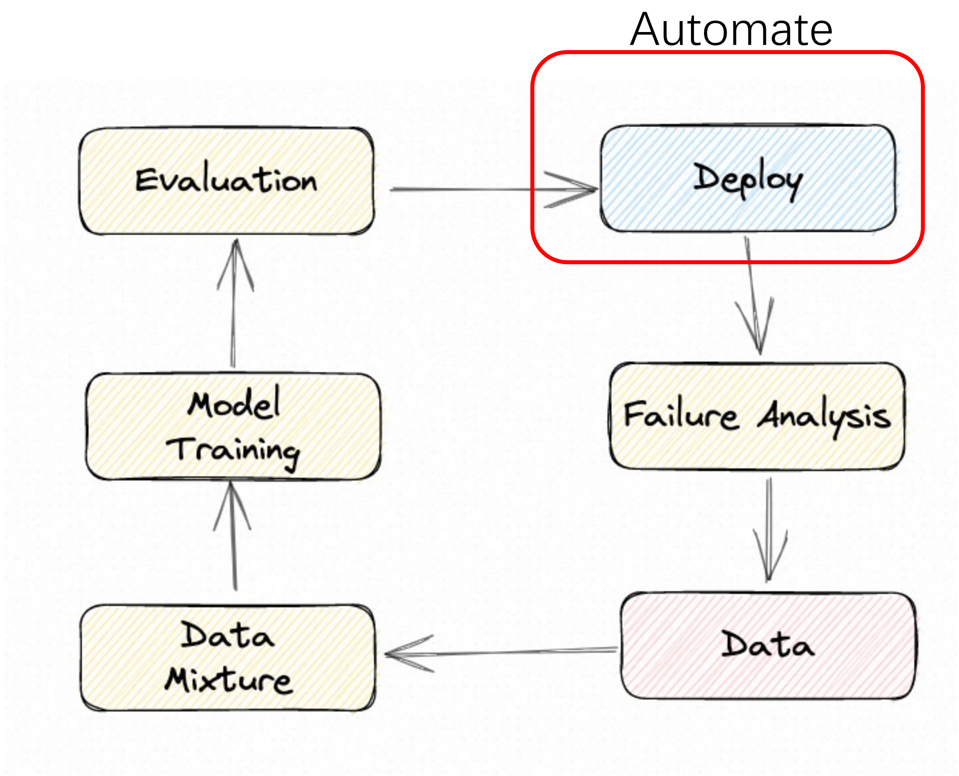




- All steps are manually done
- Cost 90 days for 1 model
  - Update an existing model
  - Does not include first design time

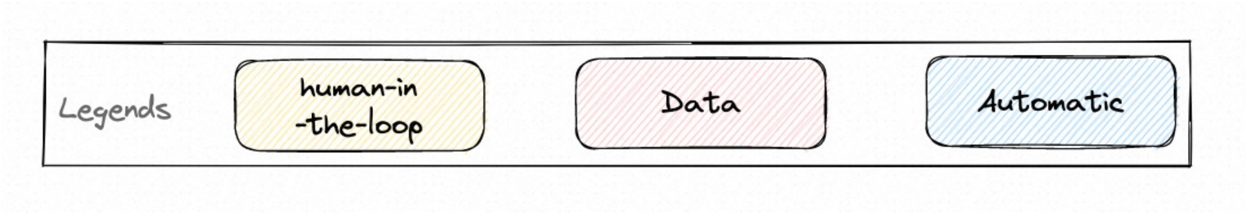
Step 1: Automatic deployment

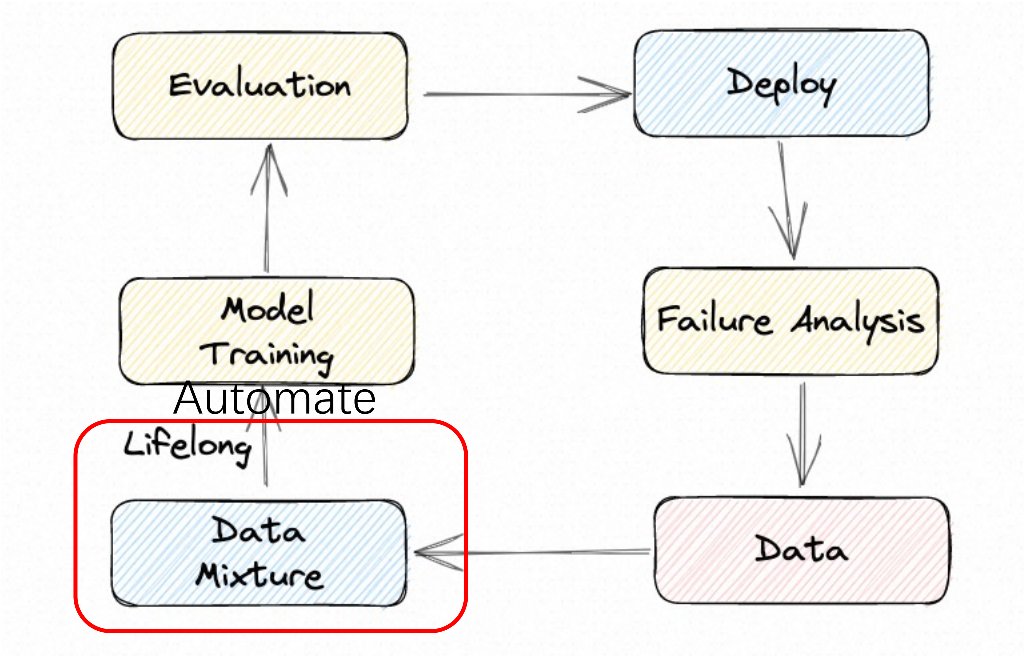




- Automation for API services
- Across 6 platforms from hard-ware deployed
- Save ~30 days

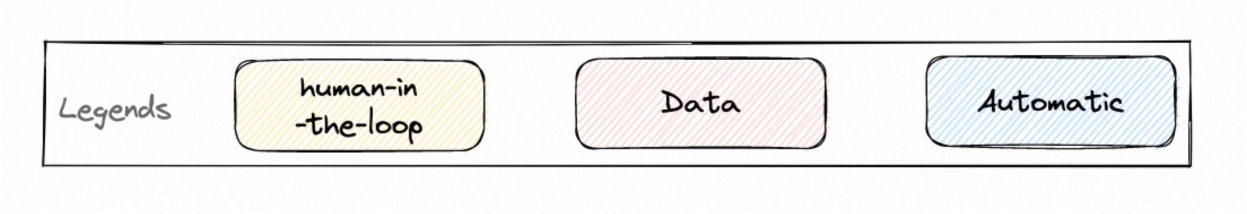
Step 2: Use active learning for data mixture process

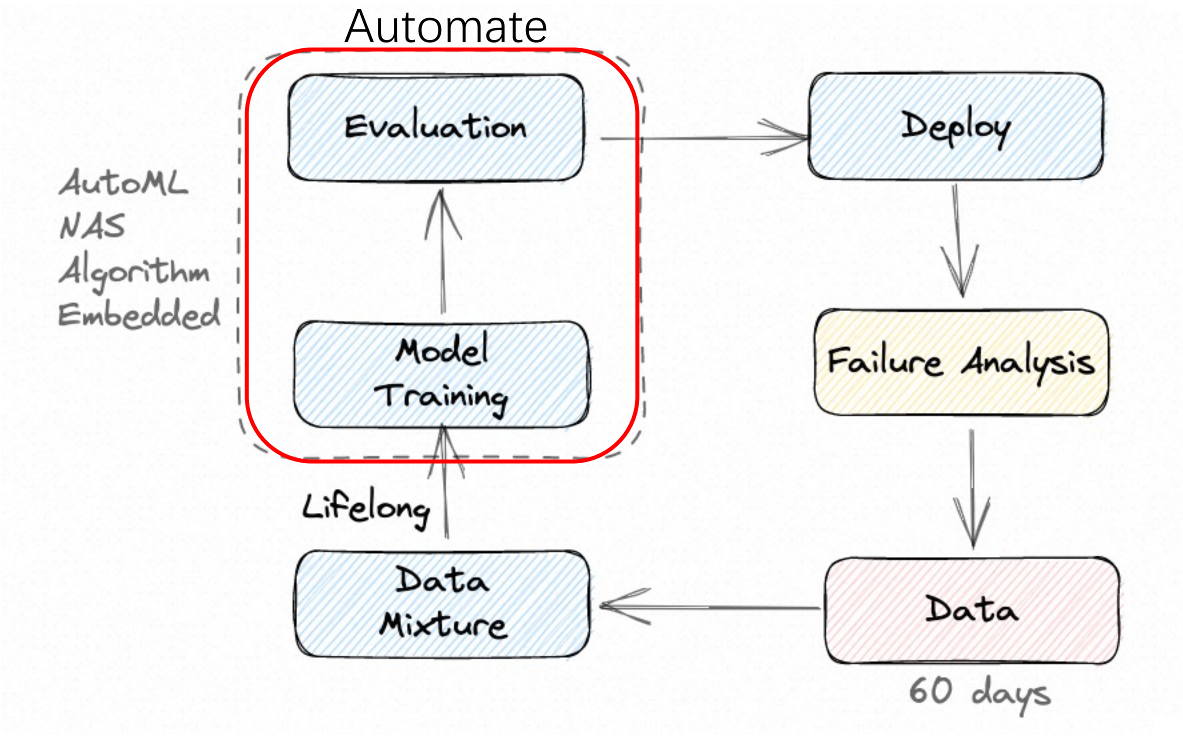




- Automatic data mixture
- Lifelong learning to train the network
- Save ~5 days
- Without performance drop

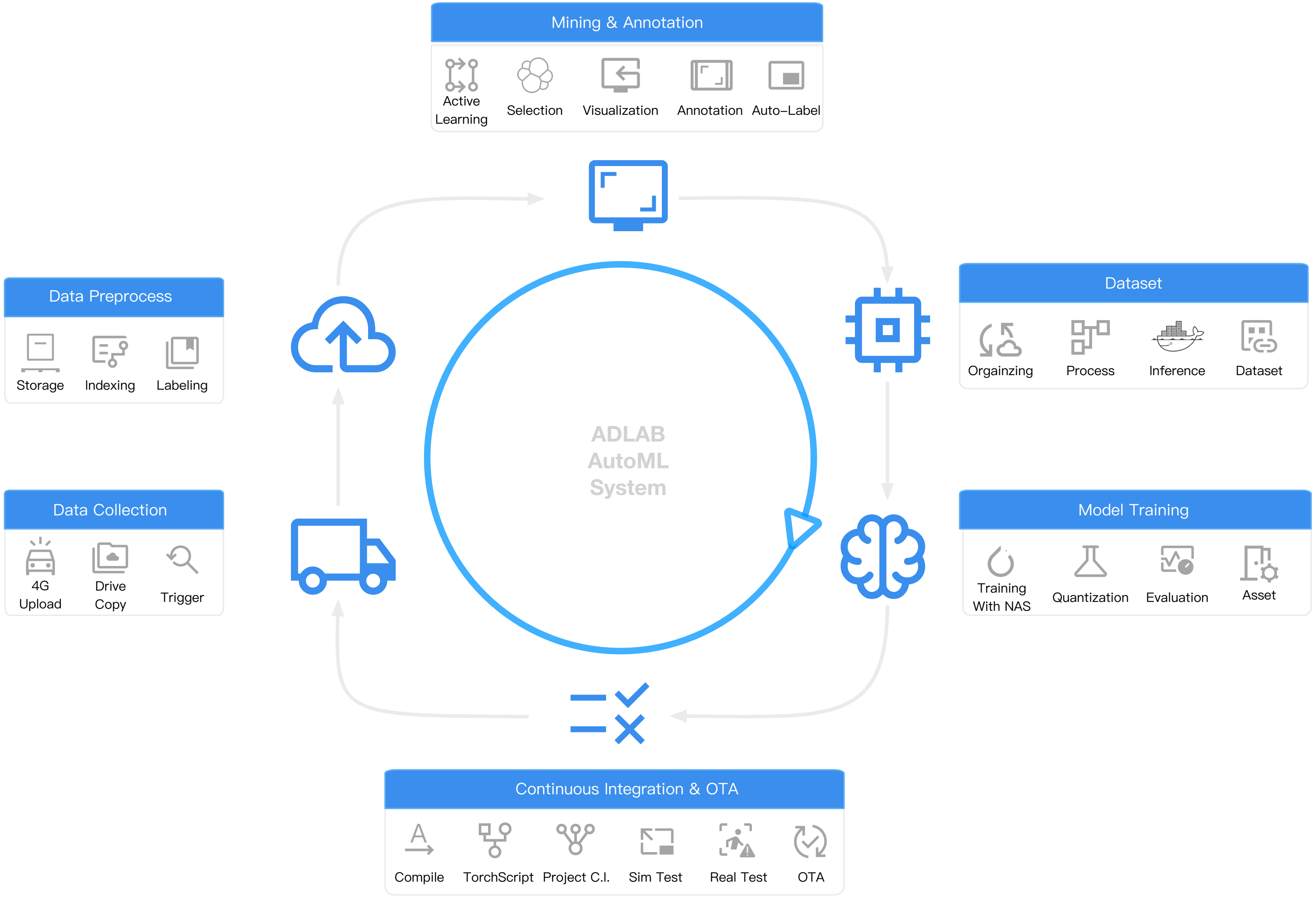
Step 3: Incorporate NAS into AutoML System





- Incorporate NAS in 3D backbone
- Support quantization
- Save ~20 days
- Performance Improves ~10%

Overview of the system



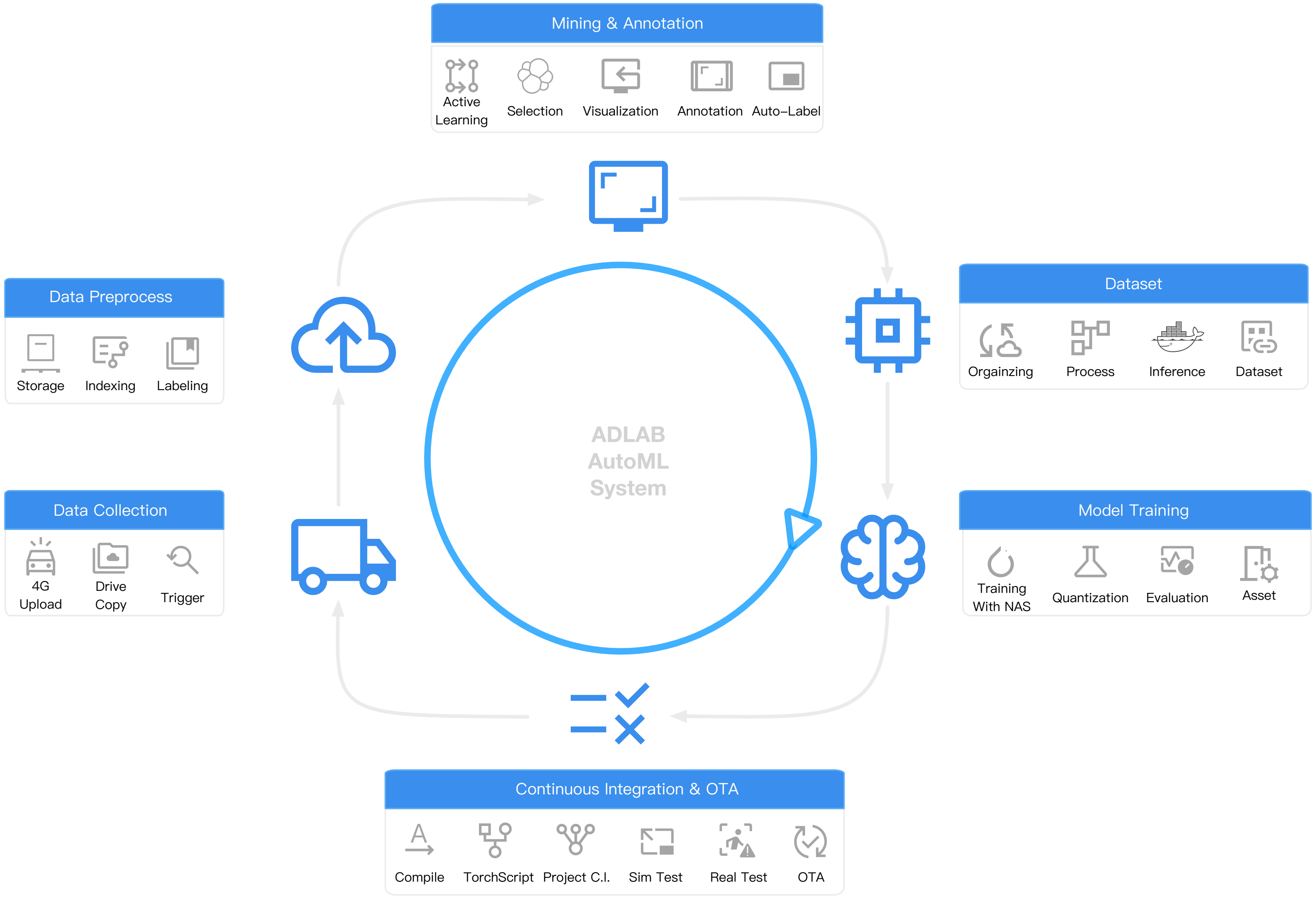
Overview of the system

Performance

- +10% mAP on object detection
- +5% mIoU on point-cloud segmentation
- Fix 150+ failures automatically

Efficiency

- Time spent: 90 → 35 (-60%)
- Manual steps: 192 → 7 (-97%)



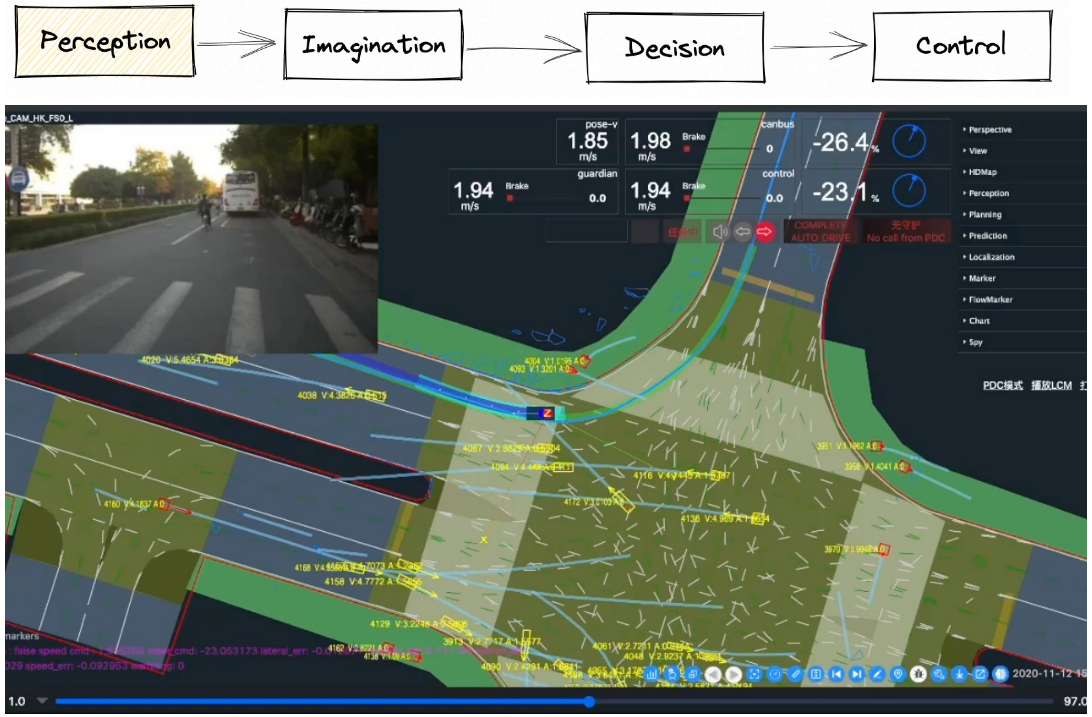
Outcome: Deployment of AutoML System V1

Carrier
Largest Autonomous Driving in logistic

200+ Cities

800+ Vehicles

rders



x 20

Before

x 1

x 1!

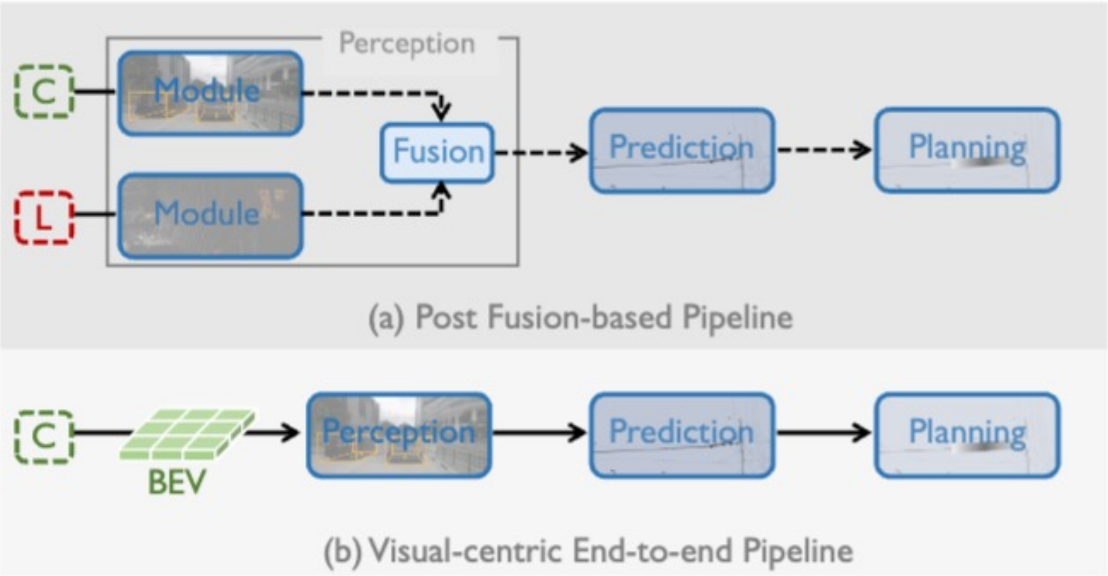
AutoML System V1

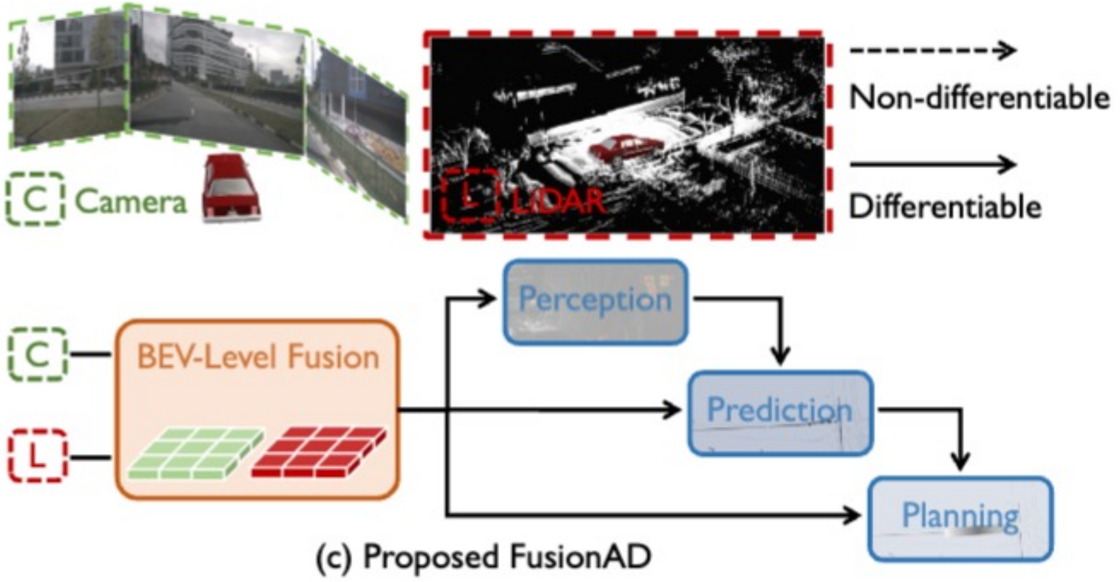

取件就约
机器人！
浙

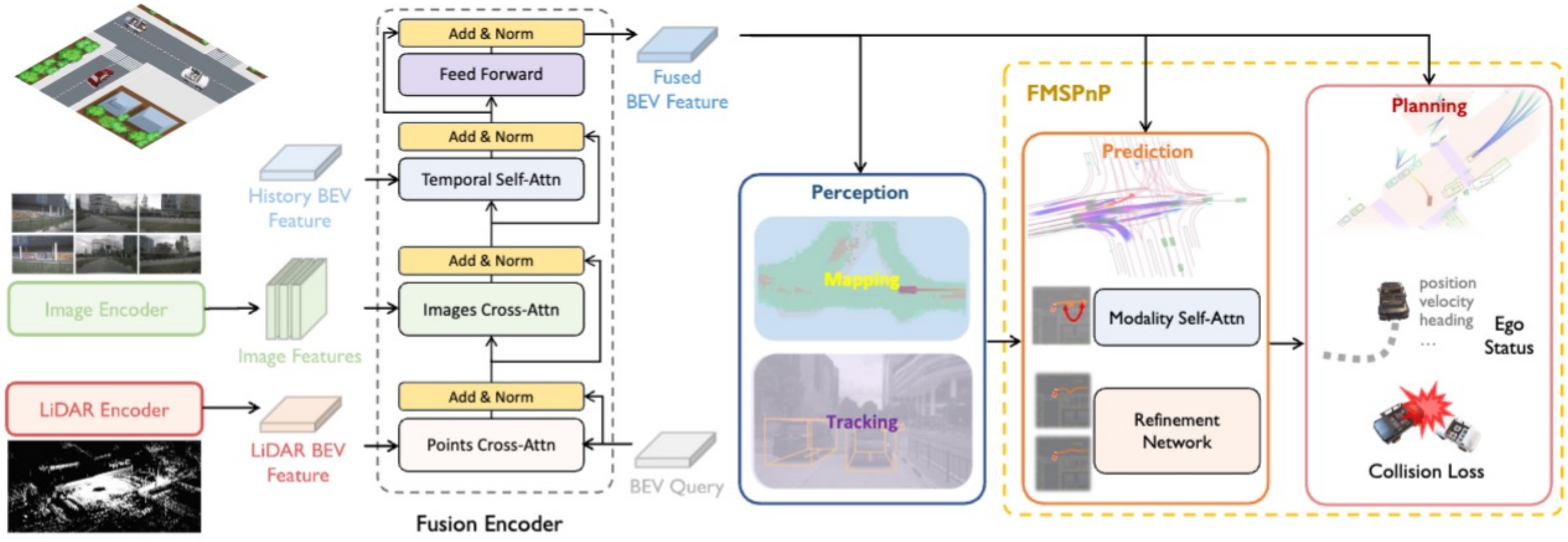
1 2 3 4 5

Conclusion
Future Work

Work in Progress: FusionAD End-to-end Autonomous Driving

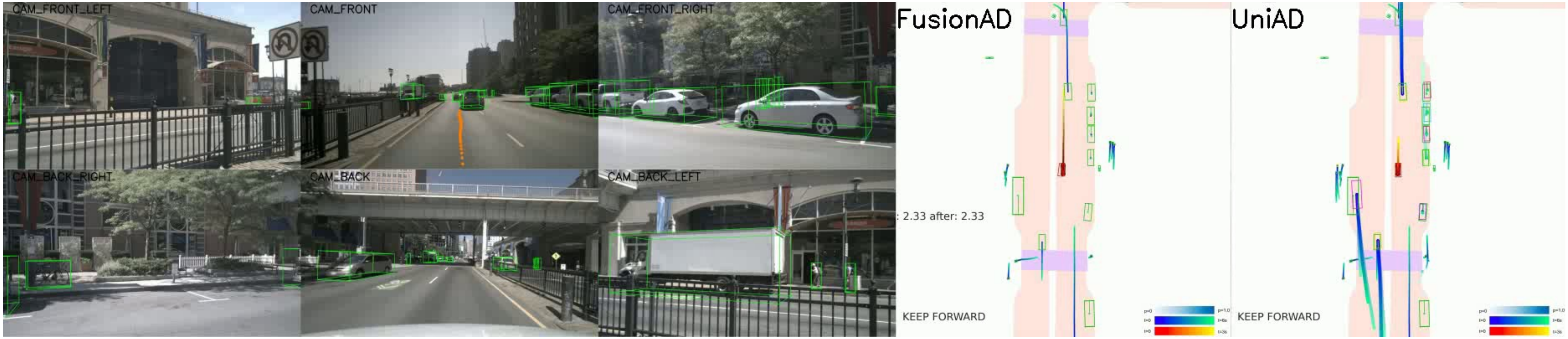






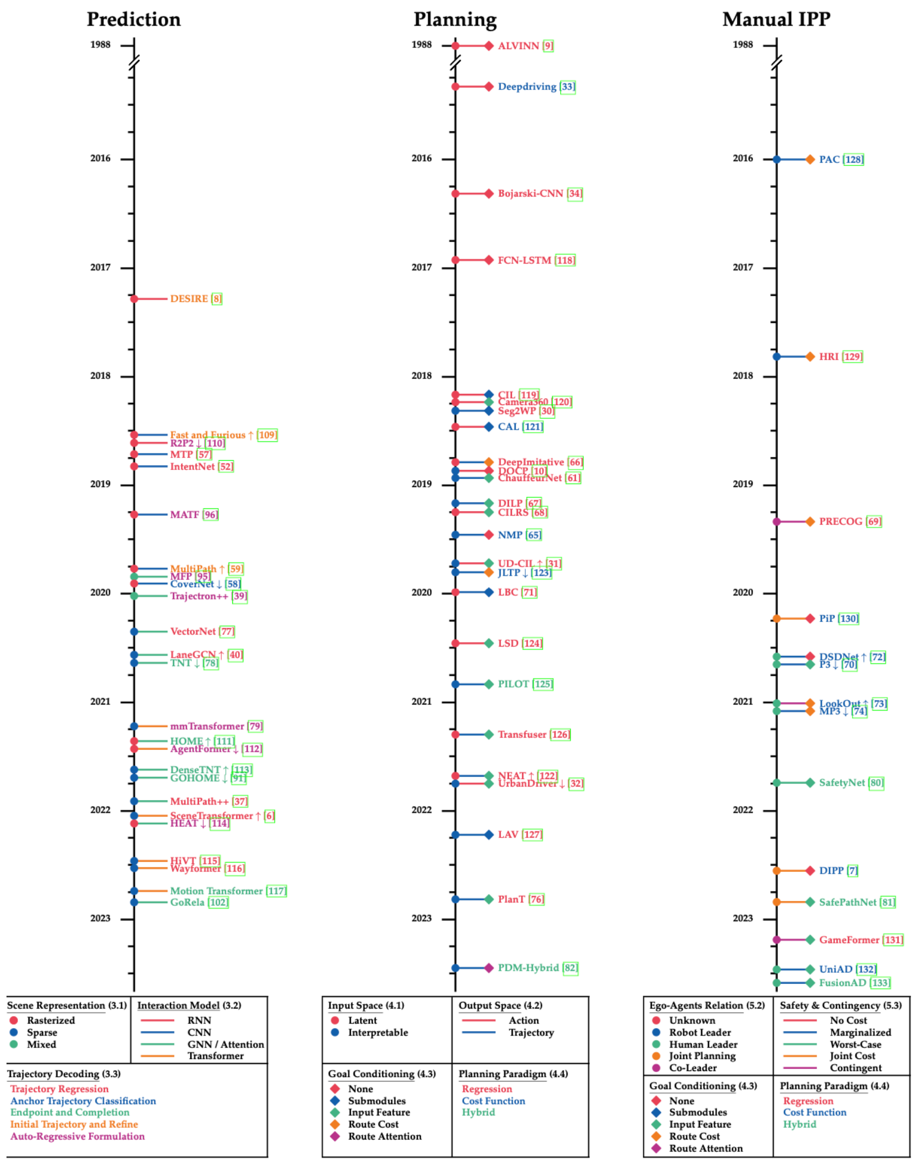
Work in Progress: FusionAD End-to-end Autonomous Driving

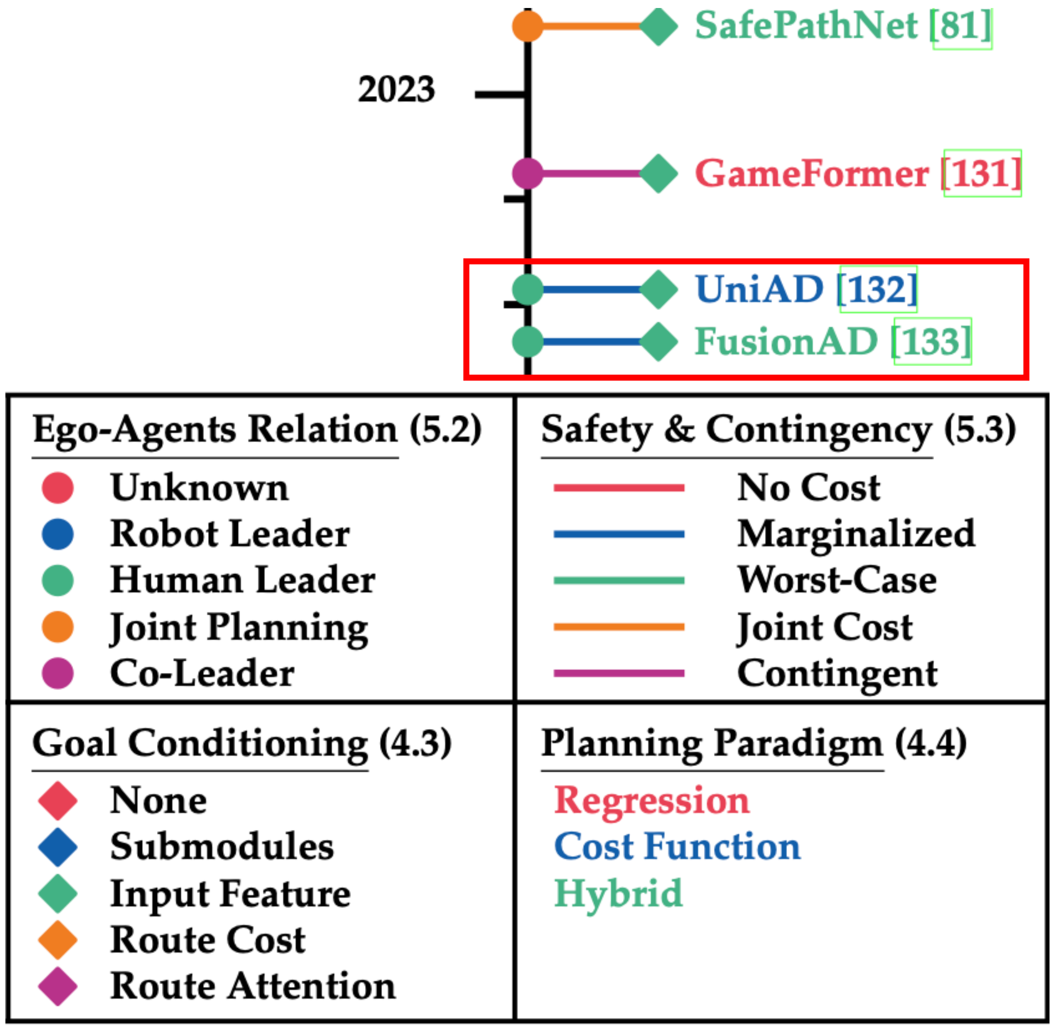
Perception of a bus. FusionAD detects the heading correctly while distorsion exists in near range, but UniAD incorrectly predicts the heading.



Prediction of U-turn. FusionAD consistantly predicts the U-turn earlier in all modes which aligns with the ground-truth trace, while UniAD still pre

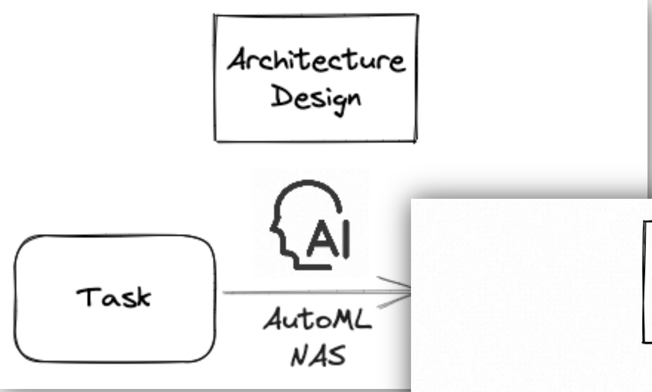
Work in Progress: FusionAD End-to-end Autonomous Driving



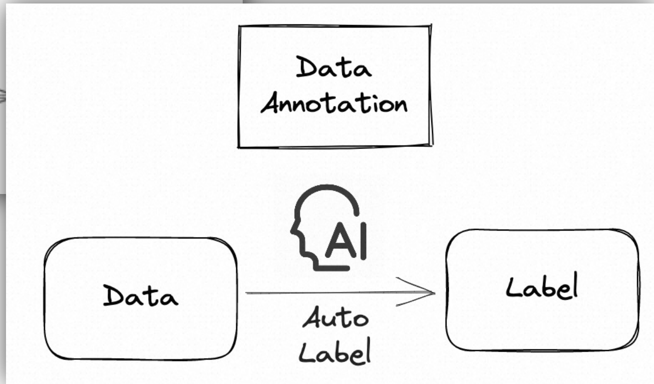


Hagedorn et al., Rethinking Integration of Prediction and Planning in Deep Learning-Based Automated Driving Systems: A Review, submitted TPAMI 2023

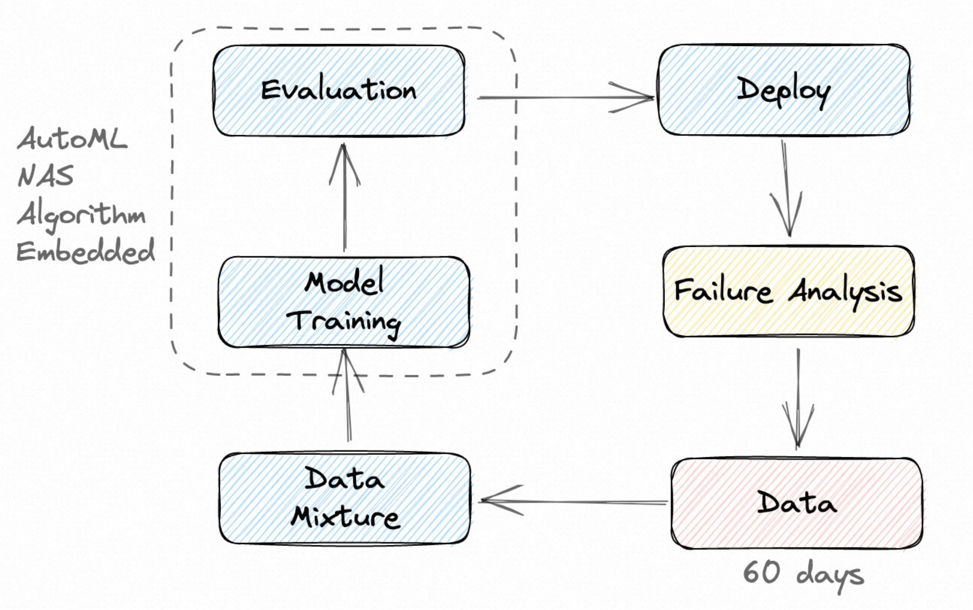
Limitation of supervised learning with given dataset



AutoML



Perception



AutoML System V1

- Assumption: Collected data contains **all** sufficient information!
- Is it really true?
  - What if we see a case **never exists** in any collected data?

Challenge: Perception Inevitably Fails when Lacking 3D Data

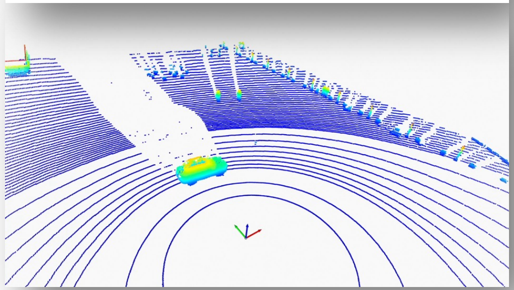
Work in Progress: Imagination via 3D Data Generation

Imagination
Data synthesis

Long way to go

- Background synthesis via segmentation mask control

- Control 3D Object in 2D Annotation

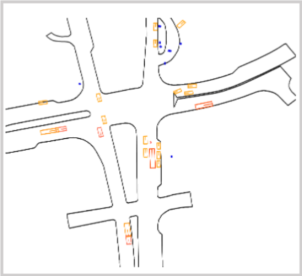


- One of the first LiDAR Simulator without reconstruction
  - LiDAR-NeRF

BEVControl: Accurately Controlling Street-view Elements with Multi-perspective Consistency via BEV Sketch Layout

Kairui Yang[1*] Enhui Ma[1*] Jibin Peng[1] Qing Guo[2] Di Lin[1†] Kaicheng Yu[3]

[1]Tianjin University [2]IHPC and CFAR, Agency for Science, Technology and Research, Singapore [3]Westlake University



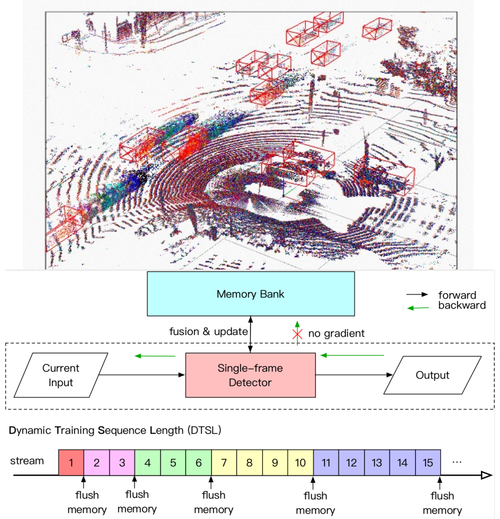
天津大学 Tianjin University

Centre for Frontier AI Research CFAR

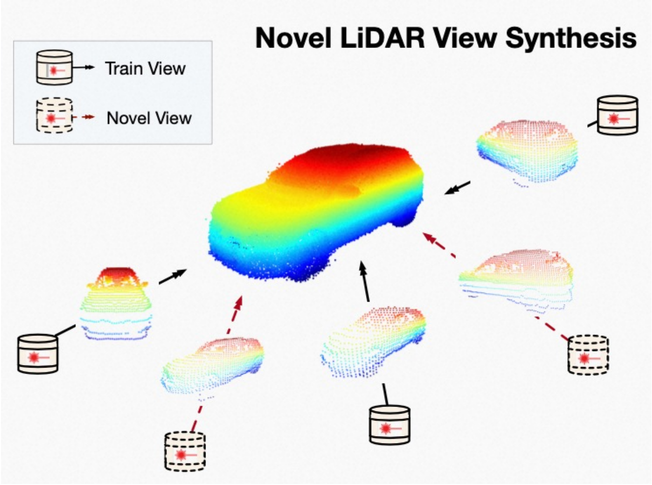
西湖大學 WESTLAKE UNIVERSITY

Other Work in 3D Perception

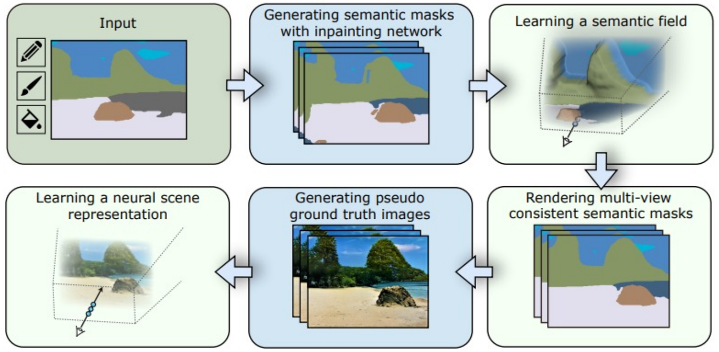
3D Backbone design



Sensor Simulation



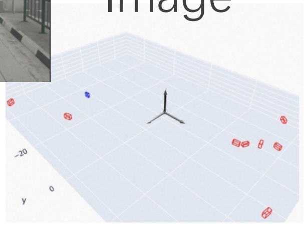
Scene Editing



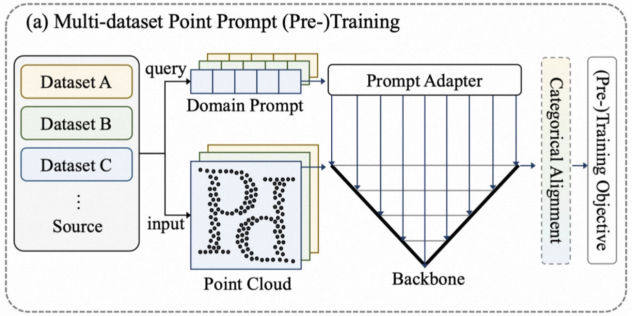
Optimizing 3D neural fields from a single semantic mask

Open World Tasks

Text
LiDAR
Image



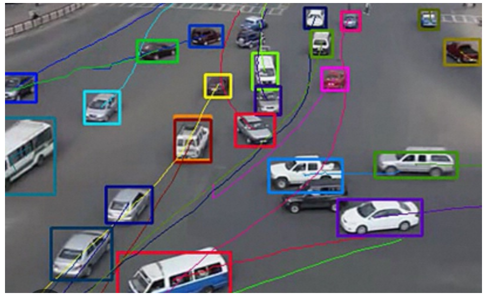
Cross-dataset pretraining



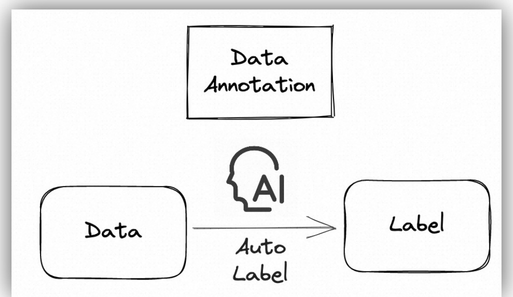
POINTCEPT
Point Cloud Perception Codebase

LLM Application

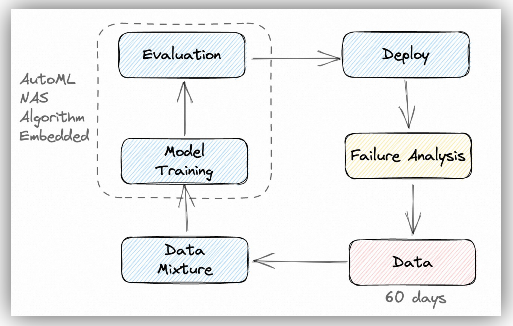
LLM +
SAM +
Tracking



Summary



Perception



Data Synthesis

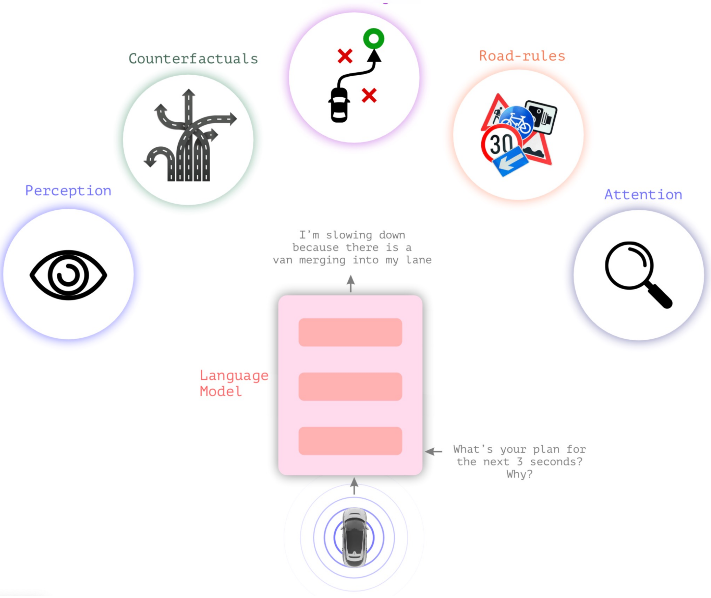
What we learn from the company:
Research never ends! Engineering approaches can never be enough to resolve long-tail issue

- BEVFusion is the first robust framework to sensor failures
- Improves +30 mAP on various settings v.s. SoTA
- Large impact in/outside Alibaba
- FusionAD as next step towards end-to-end AD system

- First differentiable LiDAR Renderer
- Diffusion methods for images synthesis
- Future: Diffusion for multi-modality output ?

What‘s Next?

- From Object-Centric Understanding
- Towards Scene-level compositional understanding
- LLM as a general understanding module
- Encode traffic rules into Autonomous Driving



Wayve Lingo-1

Thanks for all of my team members and collaborators!

- Supervised Students

Tingting Liang (Advisor: Yongtao Wang) *PhD Student, Peking University*
- *Topic: Towards robust camera-lidar fusion framework for 3D detection. Incoming research engineer at Alibaba Group*

Tao Tang (Advisor: Xiaodan Liang) *PhD Student, Sun Yet-sen University*
- *Topic: Towards generic 3D understanding via LiDAR point cloud simulation*

Yixing Liao (Advisor: Hengshuang Zhao) *PhD Student, University of Hong Kong*
- *Topic: Overcoming the domain gap via LiDAR point cloud translation with implicit fields*

Xiaoyang Wu (Advisor: Hengshuang Zhao) *PhD Student, University of Hong Kong*
- *Topic: Point Prompt Tuning: Cross dataset 3D indoor scene understanding.*

Shangzhan Zhang (Advisor: Xiaowei Zhou) *MSc Student, Zhejiang University*
- *Topic: Painting 3D in 2D: Novel view synthesis of natural scenes*

Hu Zhang (Advisor: Xin Yu) *PostDoc, Queensland University*
- *Topic: Open-world 3D object detection with cross modality features, in preparation of NeurIPS 2023*

Bicheng Guo (Advisor: Jiming Chen) *PhD Student, Zhejiang University*
- *Topic: Detection directly from neural implicit fields.*

Sihao Lin (Advisor: Xiaojun Chang) *PhD Student, Moonash University*
- *Topic: Knowledge distillation via semantic aware transformer*

Jiqi Zhang (Advisor: Xiaodan Liang) *MSc Student, Sun Yet-sen University*
- *Topic: Self-supervised learning in point cloud perception.*

Yassine Benyahia (Advisor: Anthony Davison) *MSc Student, EPFL*
- *Topic: Overcoming multi-model forgetting in neural architecture search*

Christian Sciuto (Advisor: Claudiu Musat) *MSc student, EPFL*
- *Topic: Benchmarking the robustness of neural architecture search*

- Academic collaborations


Prof. Di Lin


Dr. Mathieu Salzmann


Prof. Xiaodan Liang


Prof. Hengshuang Zhao


Prof. Xiaowei Zhou


Dr. Rene Ranftl

AutoLab: We are hiring!

Position

- Postdoc
- PhD (24 / 25 Fall)
- Research Assistant
- Remote Research Intern (6 month)

Possible Research Direction

- Pure exploration:
  Diving into the intelligence, AI Agent + Science
- Application driven:
  3D Perception, Autonomous Driving
  Solving long-tail via AI System

THANK YOU!

T. 0571-86886859
F. 0571-85271986
office@westlake.edu.cn

中国浙江省杭州市西湖区墩余路600号西湖大学(云谷校区)，310030
No. 600 Dunyu Road, Sandun Town, Xihu District, Hangzhou,
Zhejiang PR China, 310030