自动驾驶前沿技术探索与应用思考

商汤绝影 | SenseAuto @清华大学online

李弘扬

2022.4.29

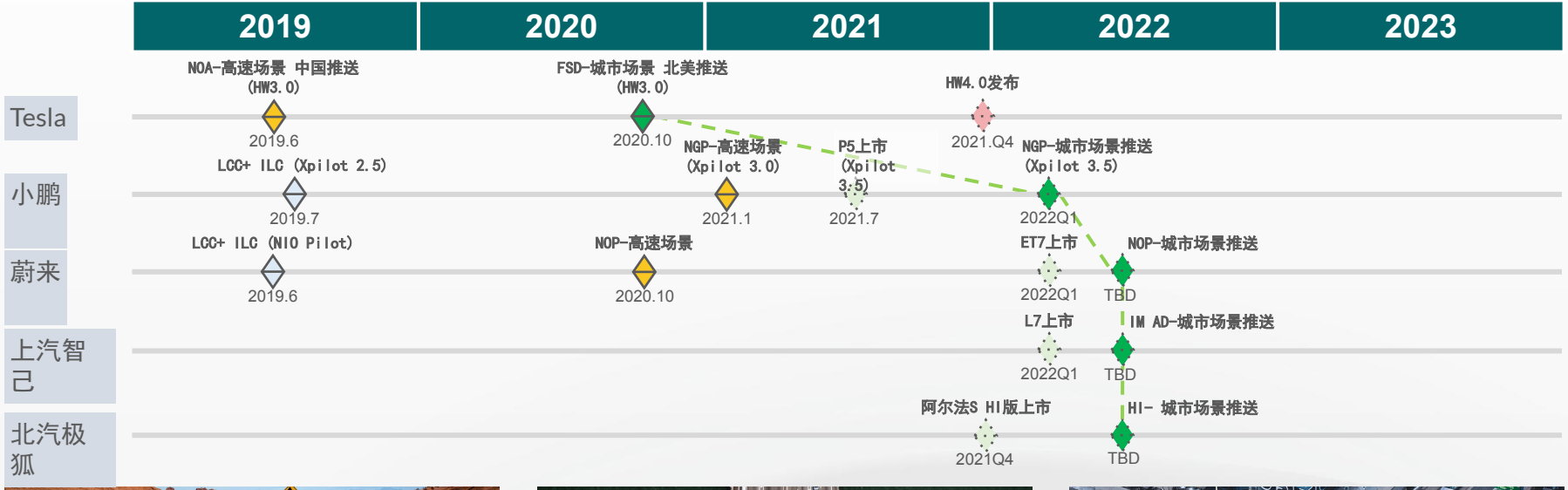
目录 | Contents

目录 | Contents

1. 自动驾驶现状
2. 自动驾驶未来：数据算法闭环体系

- 自动驾驶介绍
- 行业分析与友商对比
- 感知算法体系

行业分析 - L2+/L3 功能演进趋势



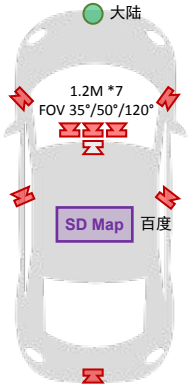
从高速场景到城市场景，从简单场景到复杂场景的导航辅助驾驶，成为L2+/L3功能主要演进趋势

系统方案对比 - 已推送

Model 3

Autopilot HW3.0 (2019/6)

NOA高速导航辅助驾驶

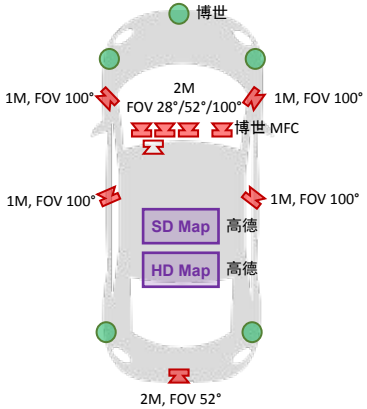


- 算法：自研
- 域控：自研
- 芯片：自研FSD芯片*2，144 Tops

P7

Xpilot 3.0 (2021/1)

NGP高速导航辅助驾驶



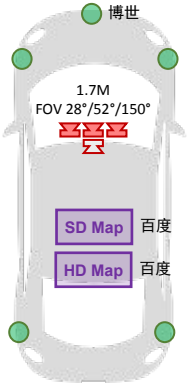
- 算法：自研
- 域控：德赛西威 IPU03
- 芯片：英伟达Xaiver (30Tops)
  + 英飞凌Aurix MCU

注：前视博世摄像头用于单独实现AEB安全功能

EC6

NIO Pilot (2020/10)

NOP高速导航辅助驾驶

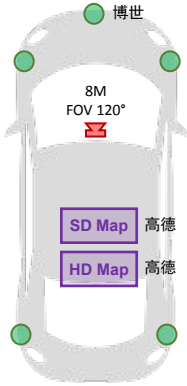


- 算法：自研
- 域控：待确认
- 芯片：Mobileye EyeQ4 (2.5Tops)
  + NXP S32V

One (2021款)

AD高级驾驶辅助系统 (计划2021/9推送)

高速导航辅助驾驶



- 算法：易航智能或自研
- 域控：待确认
- 芯片：地平线J3 *2，10Tops

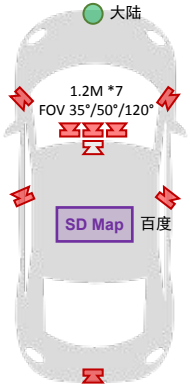
行车摄像头　DMS摄像头　毫米波雷达

系统方案对比 - 已推送

Model 3

Autopilot HW3.0 (2019/6)

NOA高速导航辅助驾驶

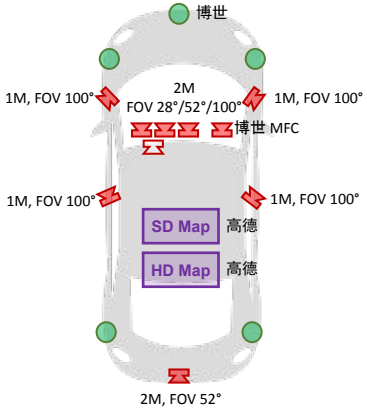


- 算法：自研

P7

Xpilot 3.0 (2021/1)

NGP高速导航辅助驾驶

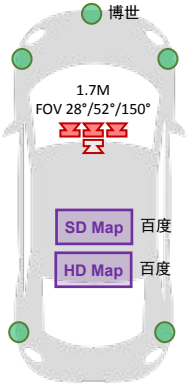


- 算法：自研

EC6

NIO Pilot (2020/10)

NOP高速导航辅助驾驶

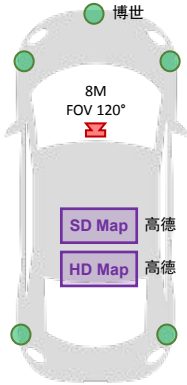


- 算法：自研

One (2021款)

AD高级驾驶辅助系统 (计划2021/9推送)

高速导航辅助驾驶



- 算法：易航智能或自研

1. 2M摄像头足以支持Highway-NOP高速公路导航辅助驾驶功能，激光雷达是非必要传感器
2. Tesla是全球L2+功能的引领者，小鹏和蔚来是国内L2+功能的引领者

行车摄像头　DMS摄像头　毫米波雷达

摄像头安装布置对比

Tesla Model 3

小鹏 P7

蔚来 EC6

理想 One（2021款）

1. Tesla和小鹏引领了智能行车周视摄像头的感知应用
2. 后视摄像头布置在后备箱处，容易脏污遮挡
3. 蔚来和理想采用5R作为360感知方案实现NOP功能

系统方案对比 - 未推送

L7

2022 Q1交付

IM AD自动驾驶（终身免订阅）

40.88万元

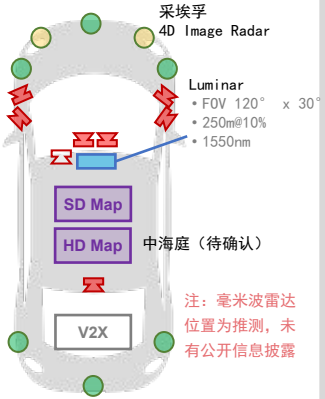
- 算法：Momenta
- 域控：创时
- 芯片：Xaiver，30Tops
- 下一代系统升级方案：
  1）摄像头像素升级到千万级；
  2）增加2个固态激光雷达；
  3）芯片升级为NVIDIA DRIVE Orin

ES33

2022 H2交付

PP-CEM高阶智驾方案

未披露



- 算法：自研
- 域控：待确认
- 芯片：NVIDIA DRIVE Orin* N，500Tops+
- PP-CEM，Pixel Point Cloud-Comprehensive Environment Model
- 4D：Range，Velocity，Azimuth，Elevation

阿尔法S

2021 Q4交付

华为HI

38.89万元/ 42.99万元

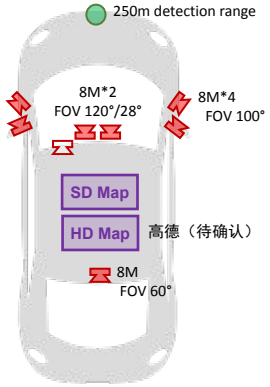
- 算法：华为
- 域控：华为MDC810
- 芯片：400Tops+
- 整车：冗余转向、制动(大陆MK C1& HBE)、电源、通讯设计

001

2021 Q4交付

NZP高速/城市自主领航

28.10万元/ 36.00万元



- 算法：Mobileye
- 域控：知行科技
- 芯片： Mobileye EyeQ5* 2，48 Tops

激光雷达　行车摄像头　DMS摄像头　毫米波雷达　4D毫米波雷达

系统方案对比 - 未推送

L7

2022 Q1交付

IM AD自动驾驶（终身免订阅）

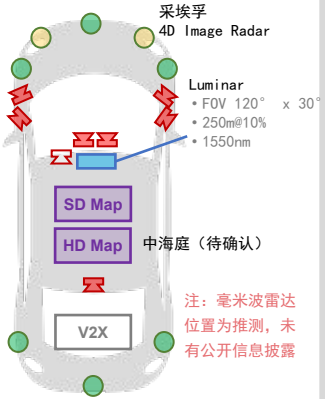
40.88万元

ES33

2022 H2交付

PP-CEM高阶智驾方案

未披露



阿尔法S

2021 Q4交付

华为HI

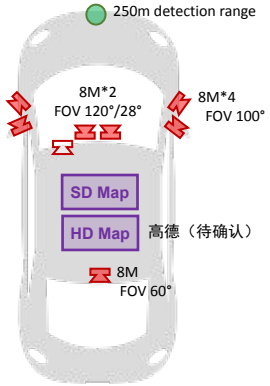
38.89万元/ 42.99万元

001

2021 Q4交付

NZP高速/城市自主领航

28.10万元/ 36.00万元



L2+系统的发展趋势：

1. 传感器配置方案趋于一致，普遍采用7*Vision/ 8*Vision + 5*Radar/ 6*Radar + 1*Lidar/ 2*Lidar/ 3*Lidar
2. 相机像素升级为5M或8M，计算平台升级为Orin/ MDC大算力平台，算法供应商不再是传统Global Tier1，新增V2X模块
3. 实现功能从高速工况导航辅助驾驶，向城市工况导航辅助驾驶演进

摄像头安装布置对比

智己 L7

R ES33

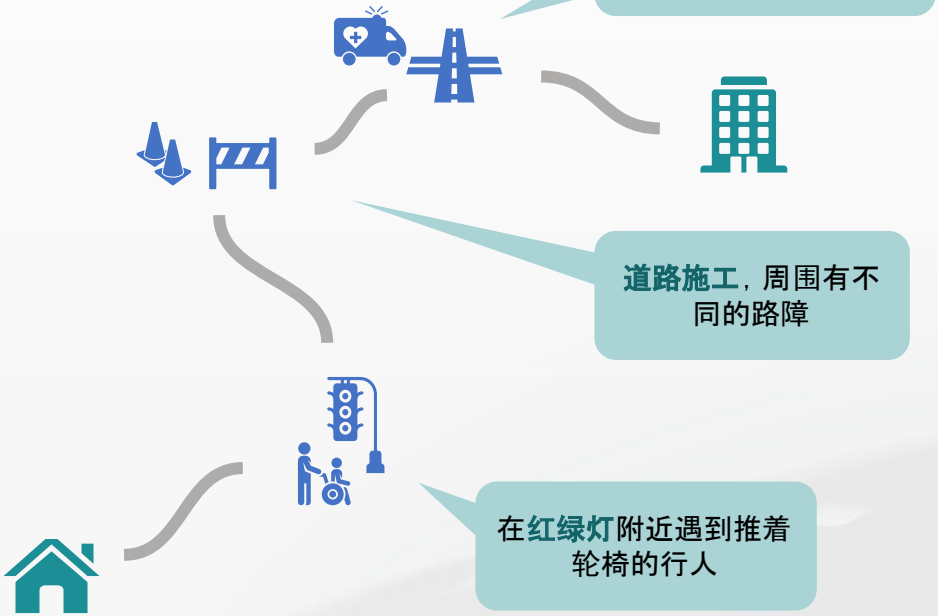
极狐阿尔法S

极氪001

主流L2+系统方案中，摄像头和激光雷达的布置位置趋同

技术需求 - L2+ - L4 功能演进对感知的技术需求

从单纯感知到场景理解，到决策AI



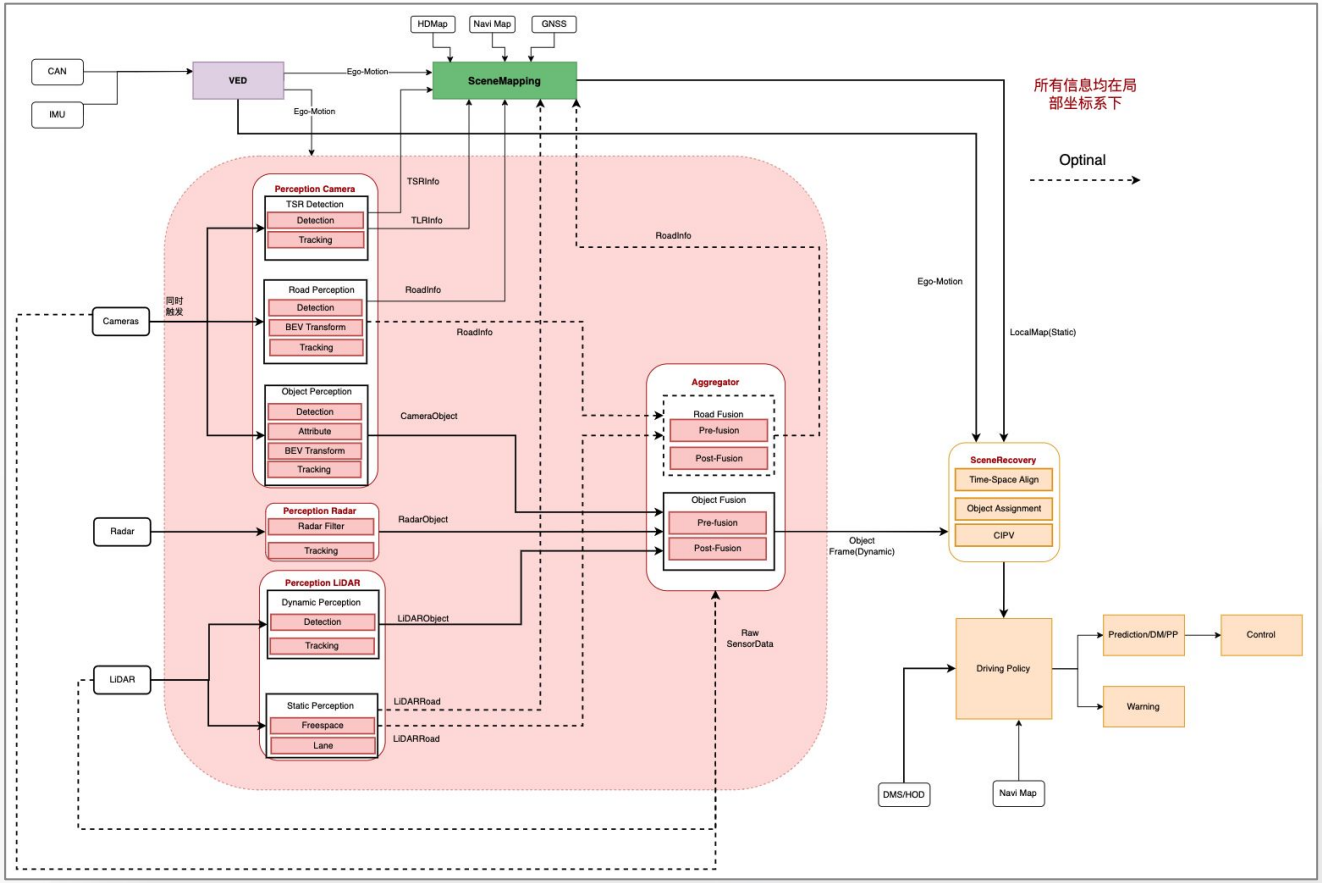
Vehicle detection

Road semantics

Traffic light recognition

L2+/L3场景覆盖度的提升，意味着对交通要素识别丰富度和长尾问题解决覆盖度的需求提升

感知算法体系



请大家观察这张图，搞清楚每个模块大致的作用；分析输入输出。

- VED
- SceneMapping
- Perception
  - Camera
  - Radar
  - LiDAR
- Aggregator
- Scene Recovery
- Driving Policy

本次talk我们仅仅涉及Perception模块中的Mono算法，nV算法；端到端感知决策一体化内容。

目录 | Contents

- 动机与概述
- 数据算法闭环核心要素分解
  - 算法举例
    - 3D Object/Lane Detection
    - nV fusion
  - 大模型训练支撑
  - 海量高质量数据支撑
    - 自动标注体系与Vidar技术
  - 算力与平台支撑
- 行业标杆案例分析

自动驾驶算法 - 迭代流程

假设解决一个长尾识别任务需要10000张训练样本

- 对长尾任务的一轮迭代至少花费2-3个月
- 一般为解决一个长尾场景，需要进行2-3轮迭代

- 假设采用10量数据采集车，每天可以采集200张有效图片，需要50个工作日完成一轮收集
- 每辆数据采集车配备司机一名，采集员一名
- 针对Corner Case需要多轮增量数据采集
- 低频场景，例如雪天车道线识别，沙尘天气识别需要更长的数据采集周期

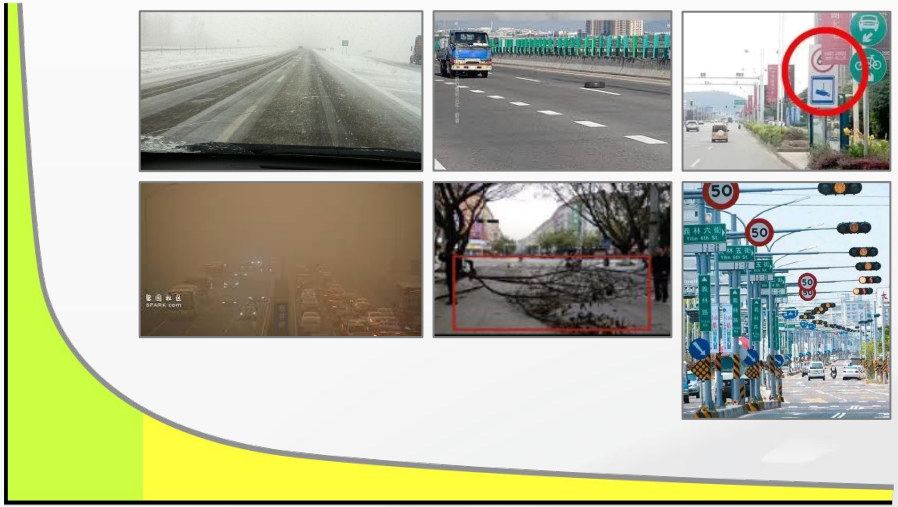
- 10000张训练样本需要做详细标注，假设每张人工标注时长1分钟，则需要167个小时，2名标注员需要10天完成标注

- 从建立模型到完成训练短则2-3天，长则1-2周

数据的采集/标注/训练需要大量时间和人力资源的投入，已 *不能* 满足L2+/L3 功能快速迭代更新需求

自动驾驶感知算法瓶颈 - 长尾分布

长尾场景检测模型



定制化长尾场景识别

- 支持定制化开发长尾场景检测模型
- 采集长尾低频场景数据，例如雪天车道线，被遮挡的交通标志牌，道路上的废弃轮胎，台风过后的挡路树枝等

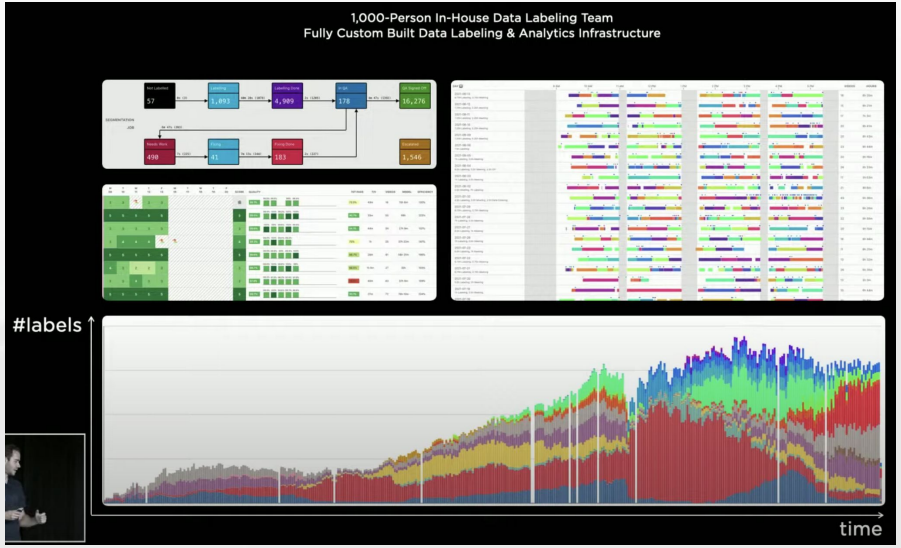
自动驾驶感知算法瓶颈 - 海量高质量数据生成

高质量数据集要素

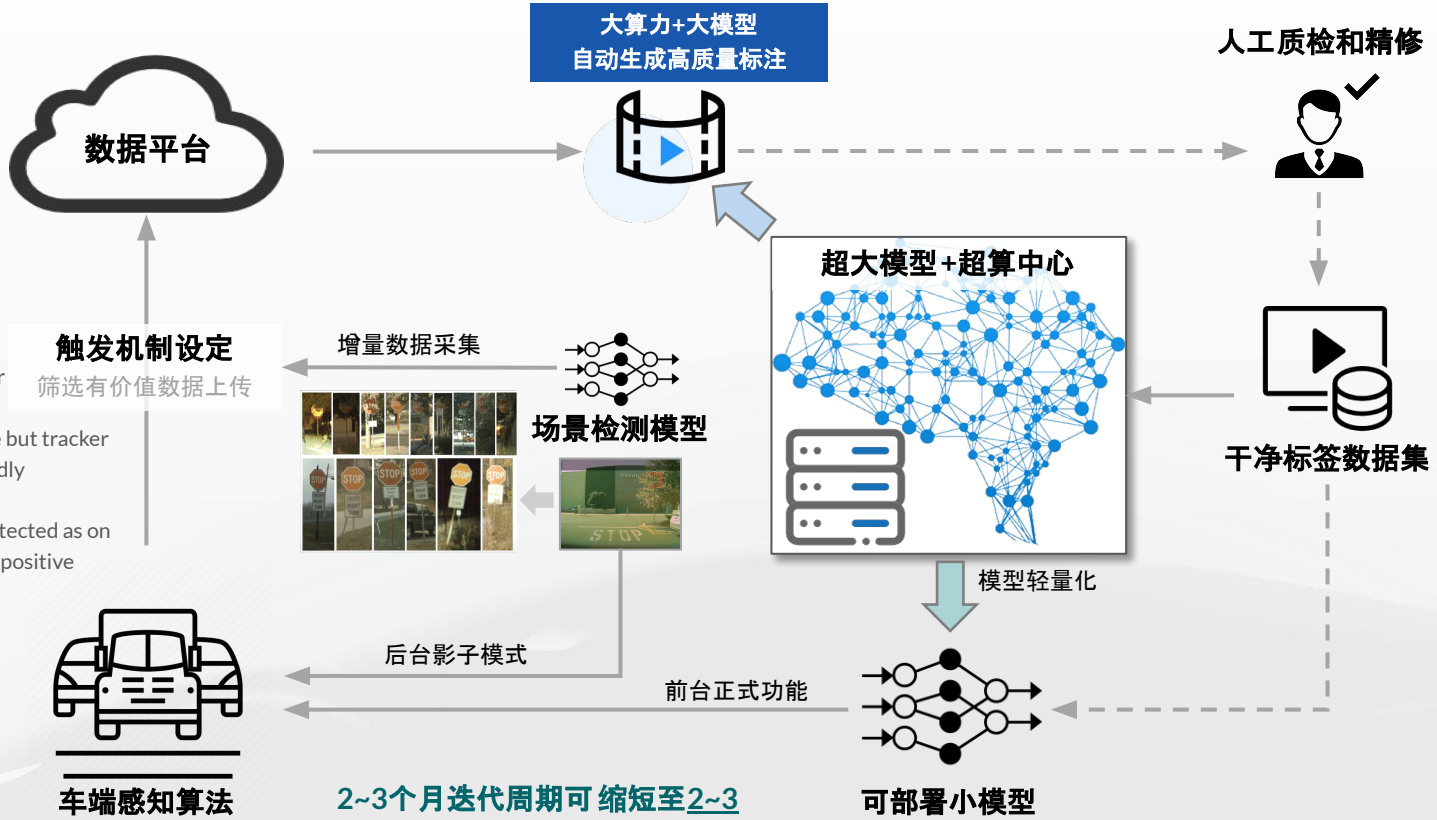
- Large-scale
- Diverse
- Clean, high-quality data (GT)

How large-scale?

- 参考 : Tesla
  - 60亿标签
  - 250w video clip
  - 1000+ 标注员



智能驾驶数据闭环解决方案



- Bounding box jitter
- Detection flicker
- Driver didn't brake but tracker thinks CIPV is rapidly decelerating
- Brake lights are detected as on but acceleration is positive
- ...

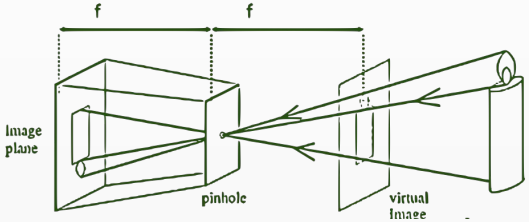
商汤具备实现智能驾驶数据闭环解决方案的全栈技术能力

数据算法闭环体系 - 算法举例

- 3D Object Detection/Lane Detection
- nV Fusion Algorithms
- 工业界与学术界区别

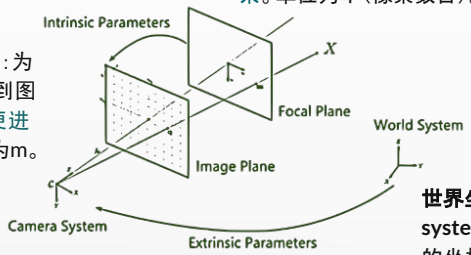
Prerequisite: 3D Vision and Camera Model

小孔成像原理



像素坐标系(pixel coordinate system)：为了描述物体成像后的像点在数字 图像上（相片）的坐 标而引入，是我 们真正从相机内读取到的信息所在的坐 标系。单位为个（像素数目）。

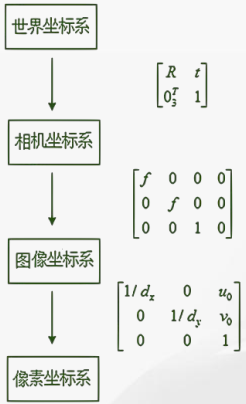
图像坐标系(image coordinate system)：为了描述成像 过程中物体从相机坐 标系到图像坐标系的投影透射关系而引入，方便进一步得到像素坐 标系下的坐标。单位为m。



世界坐标系(world coordinate system)：用户定义的三维世界的坐标系，为了描述目 标物在真实世界里的位置；单位为m。

相机坐标系(camera coordinate system)：在相机上建立的坐 标系，为了从相机的角度描述物体位置而定 义，作为沟通世界坐 标系和图像/像素坐标系的中 间一环。单位为m。

各坐标系转换关系



- u、v表示像素坐 标系中的坐标
- s表示尺度因子
- fx、fy、u0、v0、γ（由于制造误差产生的两个坐标轴偏斜参数，通常很小）**表示5个相机内参**
- **R, t表示相机外参**
- Xw、Yw、Zw（假设标定棋盘位于世界坐 标系中Zw=0的平面）表示世界坐 标系中的坐 标

$$\begin{bmatrix} \mu \\ v \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & \gamma & \mu_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

单应性矩阵 Homography Matirx

单应性变换，可以简单的理解为它用来**描述物体在世界坐 标系和像素坐 标系之间的位置映射关系**。对应的变换矩阵称为**单应性矩阵**。在左侧式子中，单应性矩阵定义为：

M为内参矩阵

$$H = s \begin{bmatrix} f_x & \gamma & \mu_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} = sM \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}$$

同时包含相机内参外参

https://zhuanlan.zhihu.com/p/54139614

3D Lane Detection

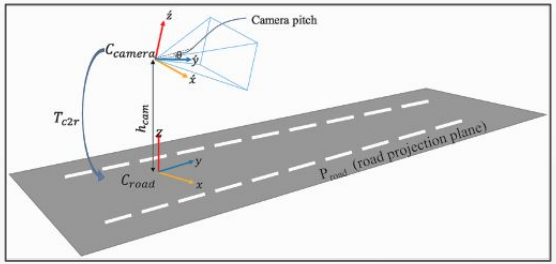
1. 从图像分割任务(u,v)到预测坐标点(x,y)

Steps:

1- 图像平面的车道线检测

2- 投影到 (水平) 地面 (IPM, inverse perspective mapping)

3- 根据车道线模型进行拟合等后处理



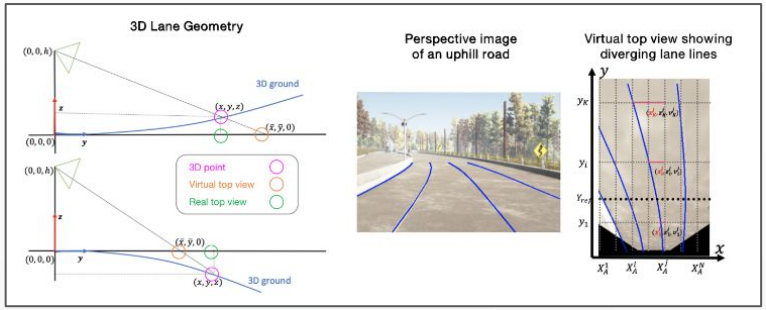
Monocular Lane Detection目前存在的问题:

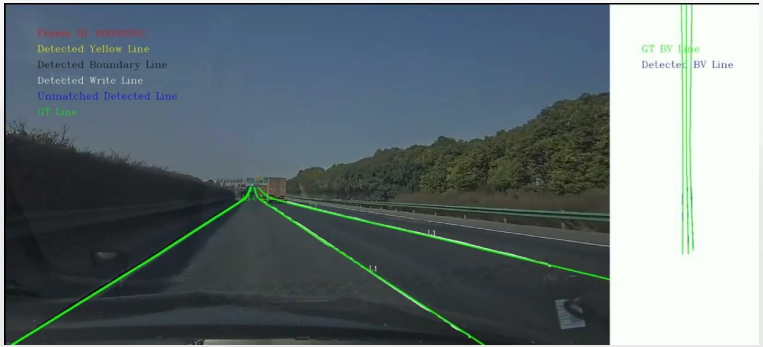
1- 依赖外参，BEV下的抖动

2- 水平路面假设在复杂场景 (上下坡等) 不成立

3- 学术数据集 (CULane/TuSimple等) 在线型/路沿等attribute上的缺失，与量产应用的差异

2. Why 3D? 从BEV(x,y)到考虑高度(x,y,z)



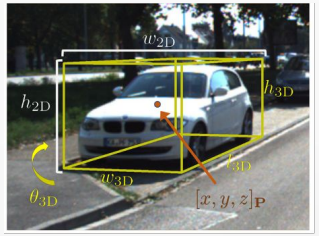


3D Object Detection

CVPR 2021
https://arxiv.org/pdf/2011.08464.pdf

任务定义

3D 物体检测是自动驾驶的关键任务，预测、意图判断、规划控制等模块均需要**3D感知信息**作为输入源。3D信息来源可以由 Lidar、Radar、Camera提供。其中Camera与Lidar、Radar等传感器相比，它提供了一个 配置简单、成本低、鲁棒性高的解决方案。



3D 任务输出信息：

- 2D bbox: (x_min, y_min) (x_max, y_max)
- Label: C
- Dimension: L,W,H
- Location: X,Y,Z
- Rotation: θ

CaDDN: Categorical Depth Distribution Network

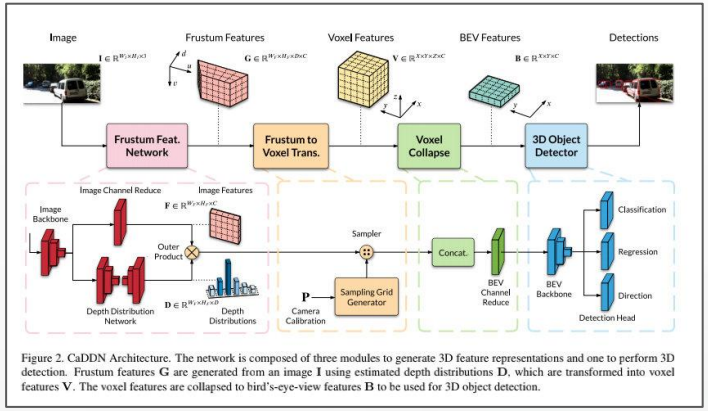


Figure 2. CaDDN Architecture. The network is composed of three modules to generate 3D feature representations and one to perform 3D detection. Frustum features $\mathbf{G}$ are generated from an image $\mathbf{I}$ using estimated depth distributions $\mathbf{D}$, which are transformed into voxel features $\mathbf{V}$. The voxel features are collapsed to bird's-eye-view features $\mathbf{B}$ to be used for 3D object detection.

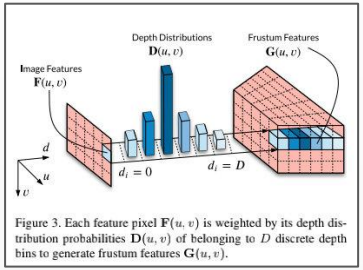
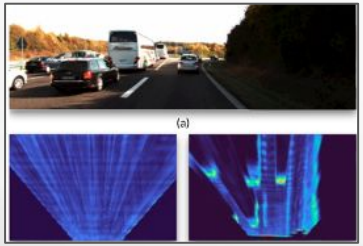


Figure 3. Each feature pixel $\mathbf{F}(u, v)$ is weighted by its depth distribution probabilities $\mathbf{D}(u, v)$ of belonging to D discrete depth bins to generate frustum features $\mathbf{G}(u, v)$.



- 独立的分支 预测深度图
- 将feature直接转到BEV下，预测3Dbox

Key Take-aways

- 增加多种图像平面的检测点，用于辅助提升3D检测的性能
- 增加2D和3D的一致性约束，有助于模型性能提升
- 深度信息对3D目标检测性能提升明显，对截断目标也有一定帮助
- BEV视角下的3D检测能将任务简化，对远近目标不一致问题又较大的优化
- 利用模型估算相机外参数，可以提升模型的泛化性能

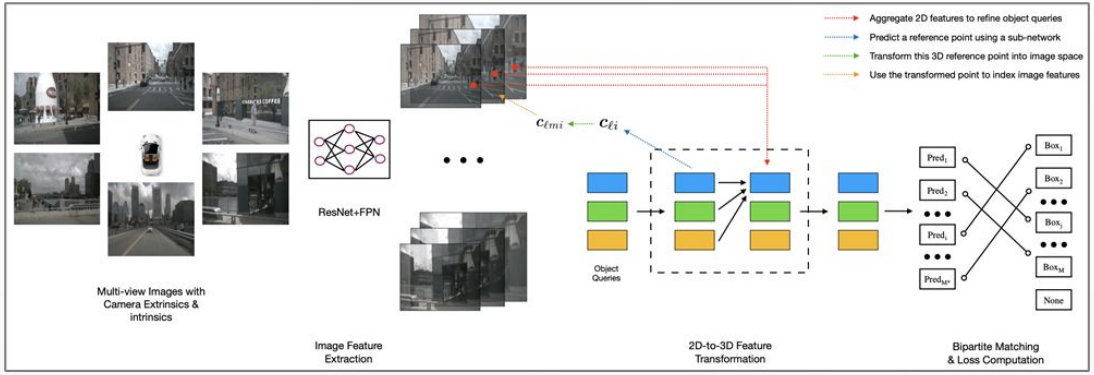
BEVFormer工作介绍

Why BEV (Bird's-eye-view)?

- 将不同视角进行统一与表征，易于信息聚合
- 没有图像视角下的尺度(scale)和遮挡(occlusion)问题

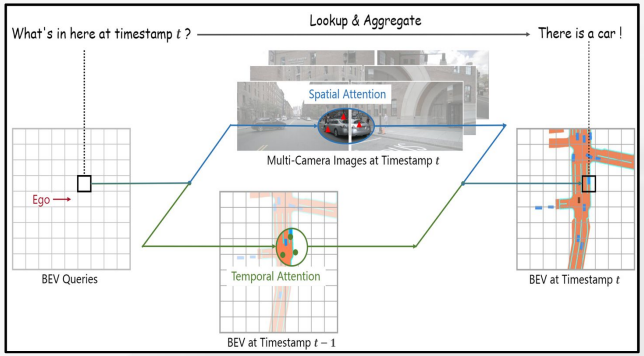
基于BEV的相关工作

- 学术界：
  BEVFormer/DETR3D/PETR/PYVA
- 工业界：
  特斯拉/地平线/毫末智行

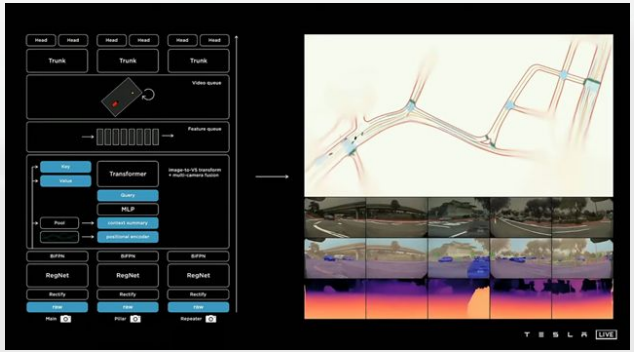


DETR3D

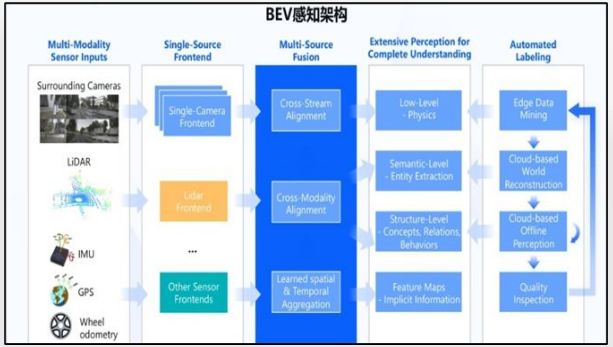
BEV感知逐渐成为自动驾驶的新范式



BEVFormer



TESLA



毫末智行感知架构图

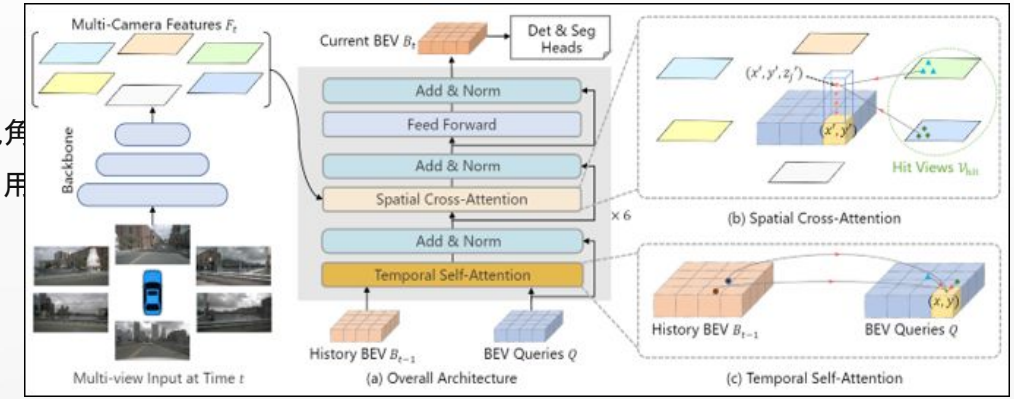
BEVFormer工作介绍

BEVFormer https://arxiv.org/pdf/2203.17270v1

基于Deformable Attention模型实现了一种融合多视角(nV)、多模态(camera/lidar)和时序特征的端到端框架，适用于多种自动驾驶感知任务

使用Transformer在BEV空间下进行时空信息融合

- **BEV queries**：统一表征
- **Spatial Cross-Attention**：用于融合多视角特征
- **Temporal Self-Attention**：用于融合时序BEV特征



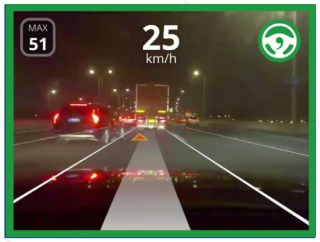
Method					Metrics								
Date	Name	Modalities	Map data	External data	mAP	mATE (m)	mASE (1-IOU)	mAOE (rad)	mAVE (m/s)	mAAE (1-acc)	NDS	PKL *	FPS (Hz)	Stats
		Camera	All	All										
2022-03-10	BEVFormer	Camera	no	yes	0.481	0.582	0.256	0.375	0.378	0.126	0.569	1.028	n/a	
2022-03-12	BEVFormer-pure	Camera	no	no	0.445	0.631	0.257	0.405	0.435	0.143	0.535	1.096	n/a	
2022-03-08	PETR-e	Camera	no	yes	0.441	0.593	0.249	0.384	0.808	0.132	0.504	1.100	n/a	
2022-03-23	Graph-DETR3D	Camera	no	yes	0.425	0.621	0.251	0.386	0.790	0.128	0.495	1.193	n/a	
2022-03-04	PolarDETR	Camera	no	yes	0.431	0.588	0.253	0.408	0.845	0.129	0.493	1.158	n/a	
2022-02-08	FudanZVG-TPD-e	Camera	no	yes	0.440	0.534	0.248	0.391	0.998	0.146	0.488	1.086	n/a	
2021-12-19	BEVDet	Camera	no	yes	0.424	0.524	0.242	0.373	0.950	0.148	0.488	1.100	n/a	
2022-03-06	SpatialDETR	Camera	no	yes	0.425	0.614	0.253	0.402	0.857	0.131	0.487	1.157	n/a	
2022-03-02	BEVDet-Beta	Camera	no	no	0.422	0.529	0.236	0.396	0.979	0.152	0.482	1.043	n/a	
2022-03-08	PETR	Camera	no	no	0.434	0.641	0.248	0.437	0.894	0.143	0.481	1.162	n/a	
2021-10-13	DETR3D	Camera	no	yes	0.412	0.641	0.255	0.394	0.845	0.133	0.479	1.207	n/a	

NuScenes 3D Object Detection Task

取得48.1 mAP和56.9 NDS，两个指标均**大幅超越**以往方法

https://github.com/zhiqi-li/BEVFormer

学术界 vs 工业界区别



Academia

检测出图中所有物体; mAP越高越好;
刷各种榜，不太考虑计算效率

Industry

只关心CIPO(危险目标); 泛化能力一定要强;
考虑算力/性能 trade off

学术界 vs 工业界区别

Philosophy 1:

相对自车，更关心**前方物体和左右车道线**，要求性能 maximize in all circumstances

Downtown

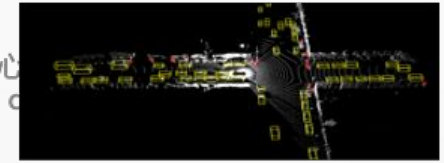
Suburban

Rainy/night

学术界 vs 工业界区别

Philosophy 1:

相对自车，更关心
maximize in all c



Philosophy 2:

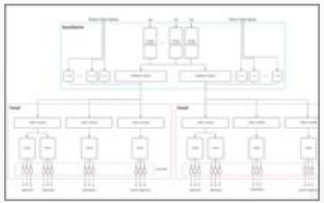
相比bounding box，更关心3D空间下(x,y,z)的**物体间位置关系**，即物体位置（**深度**）、**速度**、**加速度**、**轨迹**

Philosophy 3:

性能很重要，同时关心模型在某款芯片上部署时的**效率问题**；希望多任务学习，最大化利用芯片资源

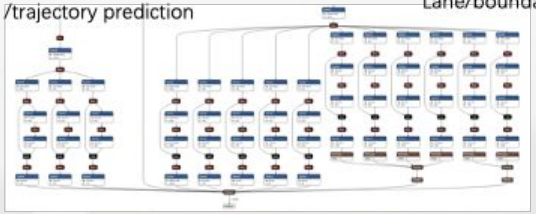
Philosophy 4:

大规模数据采集、处理、部署



Object detection
/trajectory prediction

Lane/boundary detection



TDA4vm	参数量	FLOPs	输入尺寸	输出尺寸	耗时
ResNet-10	5.416M	889.229M	3*224*224	1000*1*1	4.4ms
DenseNet_121	7.978M	3.063G	3*224*224	1000*1*1	9.7ms
mobileNet1.0	4.231M	568.740M	3*224*224	1000*1*1	4.2ms

感知下一步思路 - credit: 地平线观点

感知技术发展趋势

我们认为这些技术可能会成为达成自动驾驶的重要途径

Transformer

- 作为backbone，例如SwinTransformer
- 作为空间、时间特征融合的组件

多传感器前（中）融合

- 相比依赖人工逻辑的后融合，lidar、cam等传感器的特征级的端到端融合通过数据和计算驱动来获得更好的效果，降低对传感器同步的要求

BEV感知

- 提供了端到端的特征融合 + 视角转换
- 地平线的实践见上一讲，刘景初《上帝视角与想象力——自动驾驶感知的新范式》

（大模型的）自监督学习

- 大模型的自监督学习正显示出"智能"，例如DALL-E 2
- 但自监督预训练在下游任务的表现还有待提高

时序的端到端感知

- 结合时序特征融合的端到端检测到跟踪到预测将进一步减少系统中人工逻辑的参与

Low Level Vision

- 纯视觉自动驾驶系统的发展会依赖low成熟与应用


商汤
sensetime
绝影
SenseAuto
大模型研发基础
MOVE SMART WITH AI

大模型技术背景

- 任务通用和数据学习效率是制约当前人工智能发展的核心瓶颈问题。
- 当前的AI系统开发模式下，一个AI模型往往只擅长处理一项任务，对于新场景、小数据、新任务的通用泛化能力有限，导致面对千变万化的任务需求时，须独立开发成千上万种AI模型。
- 同时，研究人员每训练一个AI模型，都需构建标注数据集进行专项训练，并持续进行权重和参数优化。
- 这种低效的学习训练方法，导致人力、时间和资源成本居高不下，无法实现高效的模型部署。

“书生”的推出能够让业界以更低的成本获得拥有处理多种下游任务能力的AI模型，并以其强大的泛化能力支撑智慧城市、智慧医疗、自动驾驶等场景中大量小数据、零数据等样本缺失的细分和长尾场景需求。

INTERN 1.0 | Overview

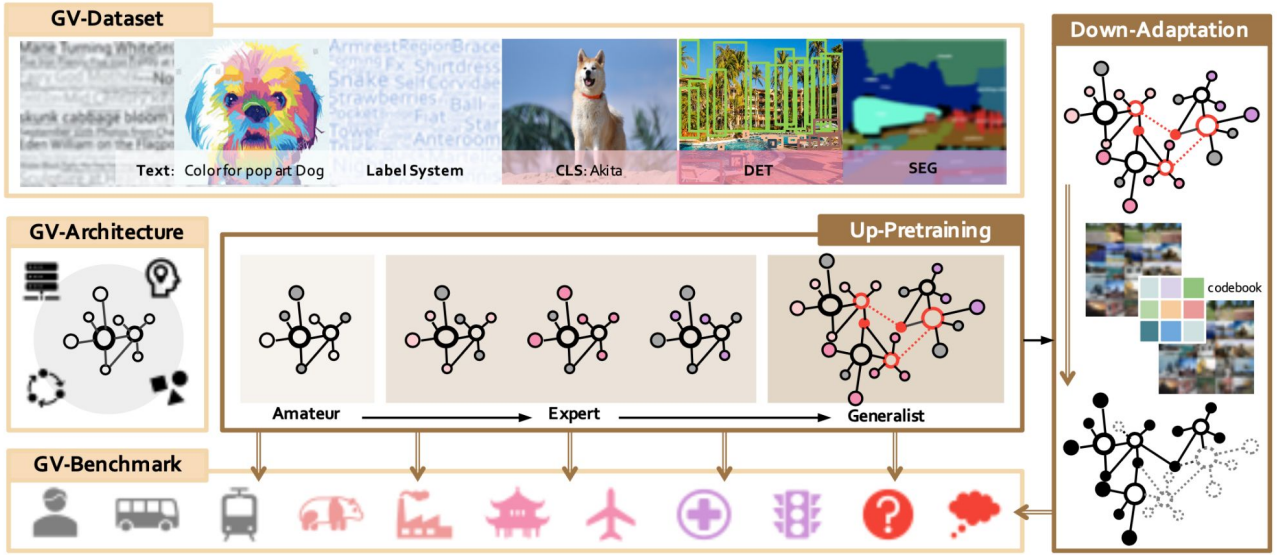


Figure 2: **Overview of INTERN.** Our complete flow of learning and evaluating a general vision model consists of three fundamental bases (*i.e.* GV-Dataset, GV-Architecture, and GV-Benchmark), a three-stage upstream pretraining scheme (*i.e.* Amateur, Expert, and Generalist), and a downstream adaptation algorithm that transfers the up-pretrained models to various downstream tasks in the benchmark. It shows that a general model (*e.g.* Generalist) with a continuous learning process exhibits stronger generalizability even on unseen tasks (shown in a red question mark).

基础模块

GV-Dataset

GV-Architecture

GV-Benchmark

上游预训练

Up-Amateur

Up-Expert

Up-Generalist

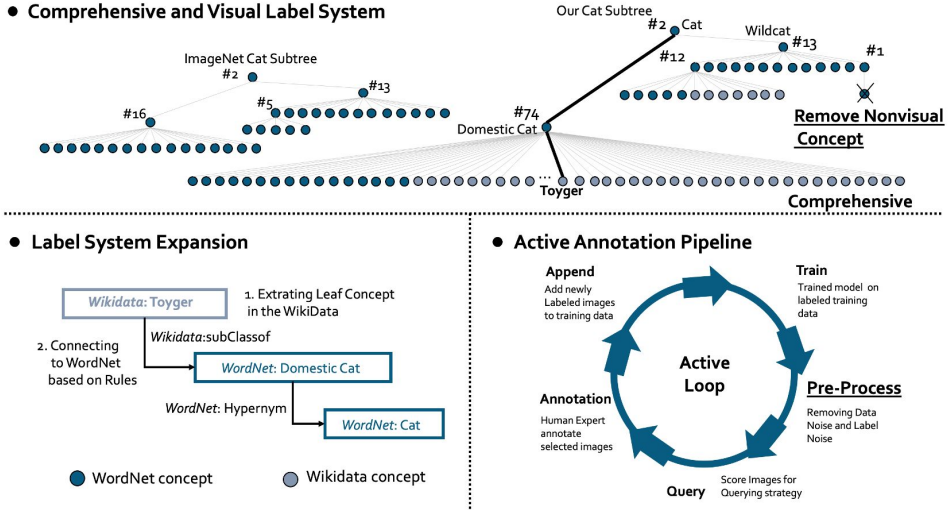
下游迁移

Down-Adaptation

10,000,000,000+
超大图像文本对数据集

100,000+
超大标签体系

39,000,000+
超大标注数据集



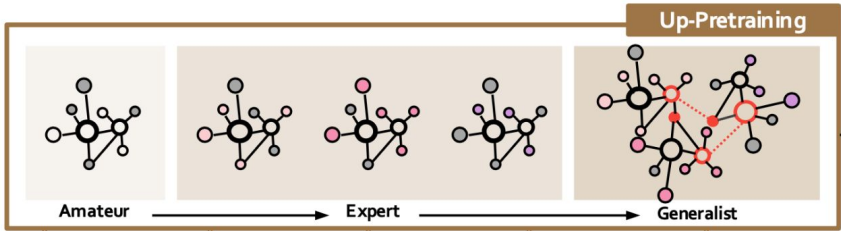
标签体系

Datasets	Concepts	Images	Labels	Open Source
YFCC-100M [37]	-	99M	99M Texts	Yes
WIT [30]	500K Queries	400M	400M Texts	No
ALIGN [24]	-	1.8B	1.8B Texts	No
GV-D-10B	**1.65M Queries**	**10B**	**10B Texts**	Partially
ImageNet-21K [14]	22K Categories	14M	14M Image-Level Labels	Yes
IG-1B [28]	17K Queries	1B	1B Hashtags	No
JFT-3B [42]	30K Categories	**3B**	3B Noisy Image-Level Labels	No
GV-D_c-36M	**119K Categories**	36M	**36M Image-Level Labels**	Yes
COCO [27]	80 Categories	118K	1M Bounding Boxes	Yes
Object365 [34]	365 Categories	609K	10M Bounding Boxes	Yes
OpenImages [26]	600 Categories	2M	15M Bounding Boxes	Yes
GV-D_d-3M	**809 Categories**	**3M**	**25M Bounding Boxes**	Yes
COCO-Stuff [4]	182 Categories	118K	Segmentation Masks	Yes
GV-D_s-143K	**334 Categories**	**143K**	Segmentation Masks	Yes

Table 5: **Summary of GV-D and other large-scale datasets for visual pretraining.** A large-scale database is a fundamental component of general vision pretraining. YFCC-100M, ImageNet-21K, COCO, Object365, OpenImages are instances of commonly used public datasets, while IG-1B, WIT, JFT-3B are proprietary ones that cannot be accessed by the community. We construct a novel data system **GV-D** with four subsets: 1) **GV-D-10B** consisting of 10 billion image-text pairs collected with 1.65 million queries; 2) **GV-D_c-36M** contains 36 million images with classification labels from our label system of 119K categories. Although GV-D_c-36M has fewer images than JFT-3B, it has the most manual and clean labels. 3) **GV-D_d-3M** composed of 3 million images with 35 million bounding boxes of 809 categories; 4) **GV-D_s-143K** with 143 thousand images and corresponding semantic segmentation masks of 334 categories.

数据集对比

INTERN 1.0 | Performance



Amateur：基于图像文本对的多模态预训练，可利用海量网络数据。

Expert：不同任务，针对性学习，有效积累任务特性，避免任务冲突。

Generalist：整合专家模型能力，使表征更通用

Model	Data Setting	Up-A	Up-E	Up-G
ResNet-50	10%	70.9	73.7	74.3
MN-B15		80.4	84.2	84.4

分阶段预训练带来持续性性能提升

Pretrain	Data Setting	CLS-AVG ↑	VOC-DET ↑
ImageNet	100%	73.0	79.5
Up-E (C)	10%	73.7	72.2
Up-E (D)	10%	53.9	87.7
Up-G (C-D)	10%	**74.3**	**87.7**

通才模型有效融合专家模型表征能力

Pretrain	CLS-AVG ↑	VOC-DET ↑	VOC-SEG ↑
Up-E (C)	73.7	72.2	57.7
Up-E (D)	53.9	87.7	62.3
Up-E (S)	47.5	75.0	71.9
Up-G (C-D)	74.3	87.7	66.2
Up-G (C-D-S)	**74.3**	**87.7**	**73.7**

更多的专家模型进一步提升通才模型的表征能力

Pretrain	Data Setting	VOC-SEG ↑	KITTI ↓
ImageNet	100%	66.0	3.09
Up-E (C)	10%	57.7	3.21
Up-E (D)	10%	62.3	3.09
Up-G (C-D)	10%	66.2	2.84
Up-G (C-D-S)	10%	**73.7**	**2.80**

对于未知任务依然具备鲁棒的迁移效果

海量高质量数据支撑

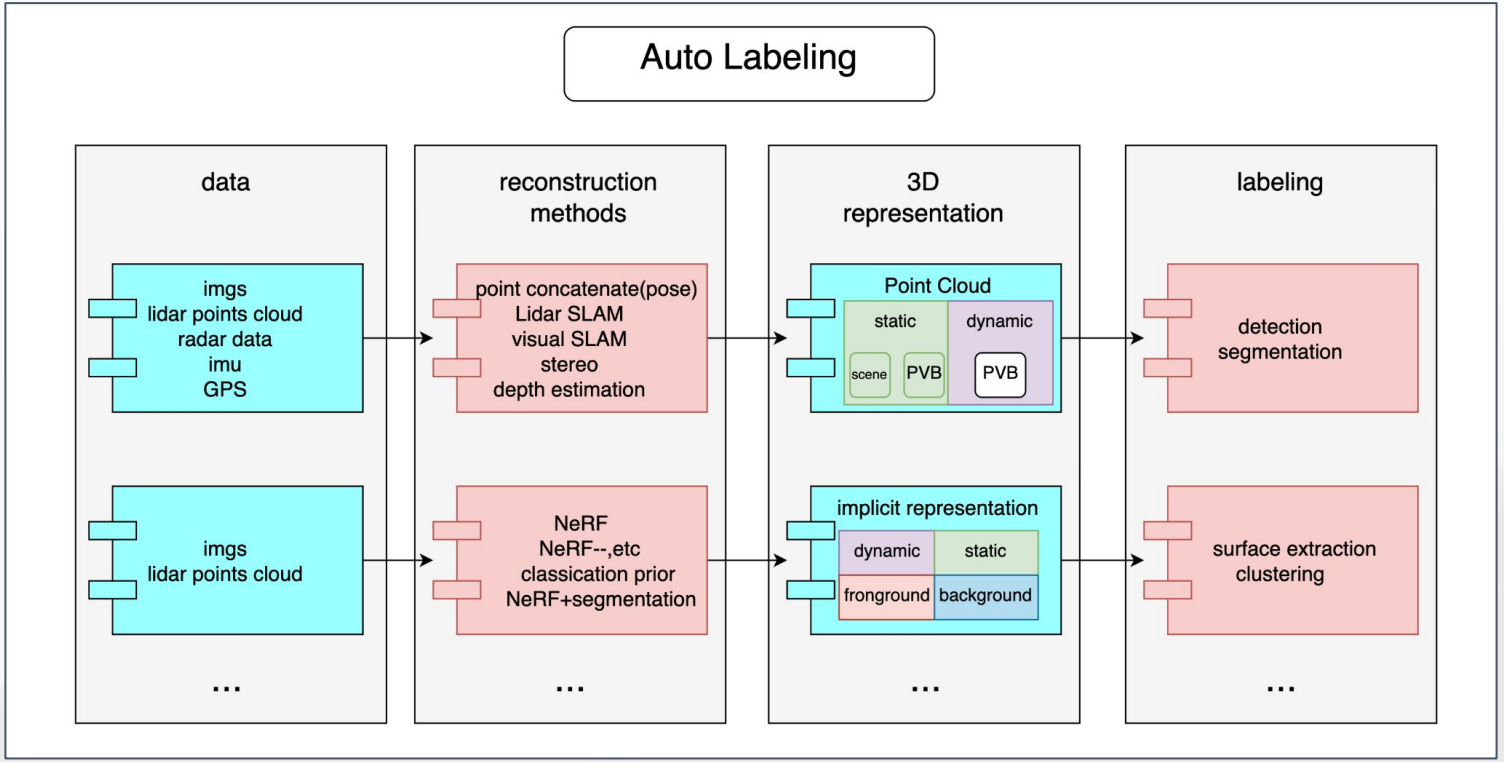
- ViDAR技术探索
- SenseMentor数据集/CVPR 2022 Competition

Auto Labelling

Life of a Clip

自动标注概述

Pipeline



自动标注 - ViDAR

Introduction

ViDAR: characteristics

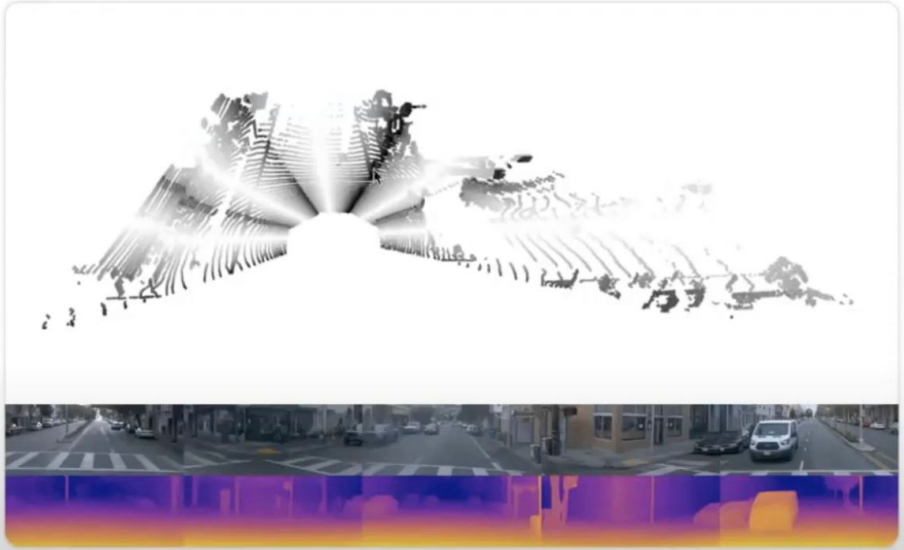
Large-scale training data and LiDAR supervision

Rolling shutter handling based on provided shutter timings

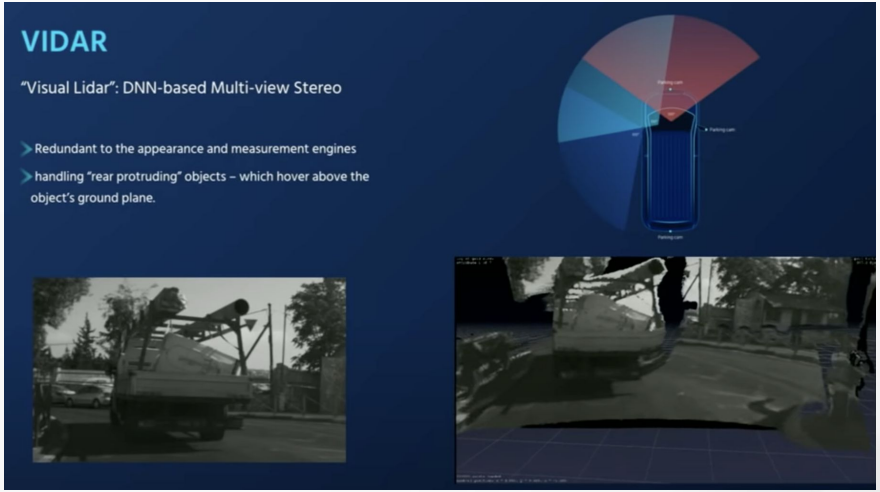
- Avoids significant displacements at higher speeds

Multi-camera support for scene-level representation

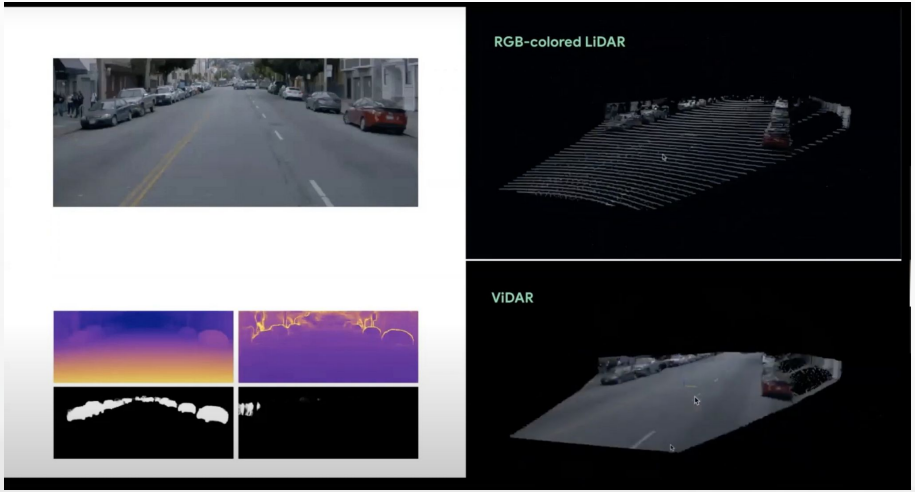
- Significant (6%+) improvements compared to single-camera model



自动标注 - ViDAR



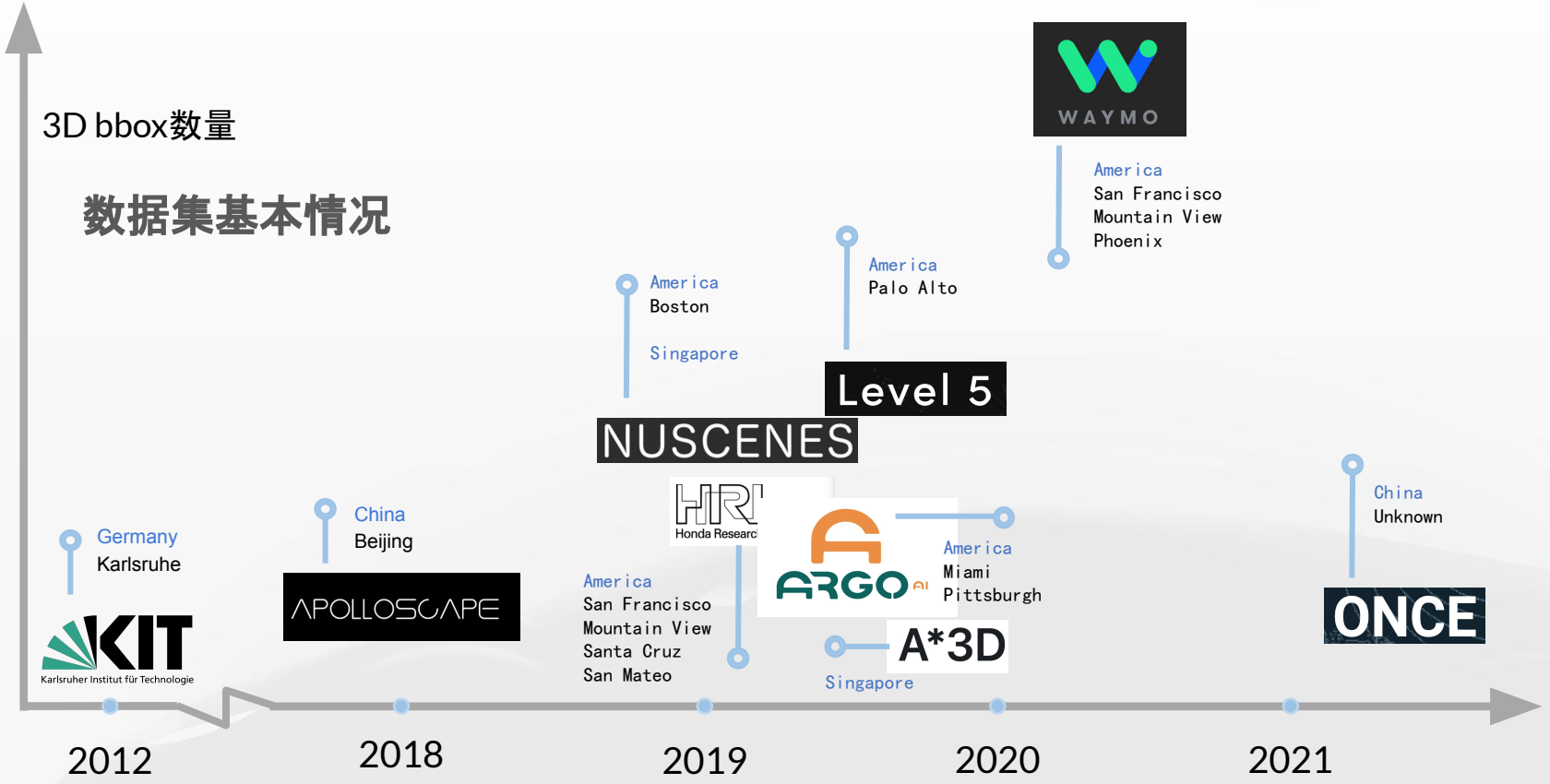
Mobileye



Waymo

OpenDrive数据集

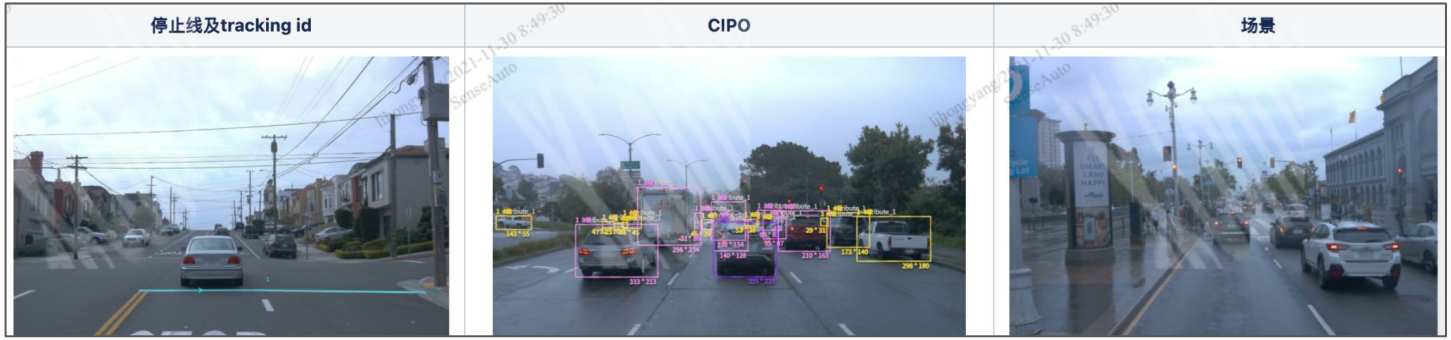
自动驾驶主流数据集

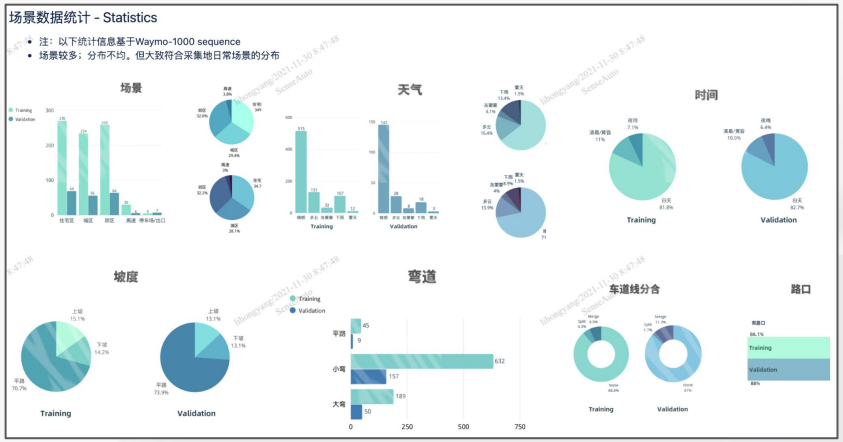


注：
1. 本图仅反应相对关系 waymo标注量远超出其他数据集
2. 随着时间增长，其他数据集也可能会有标注的补充

OpenDrive数据集

OpenLane Benchmark





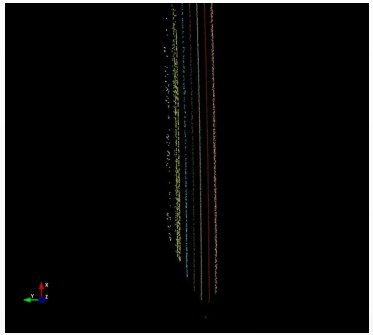
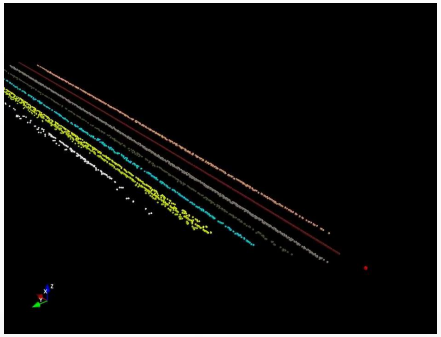
https://waymo.com/open/challenges/

Winners

Challenge 1: Motion Prediction	Method	Technical Report	Authors	Country	Affiliation
First Place	Tsinghua MARS - DenseTNT	Report	Junru Gu, Qiao Sun, Hang Zhao	China	Tsinghua University
Second Place	ReCoAt	Report	Xiaoyu Mo, Zhiyu Huang, Chen Lyu	Singapore	Nanyang Technological University
Third Place	SimpleCNNOnRaster	Report	Stepan Konev, Artsiom Sanakoyeu, Kirill Brodt	Russia, Germany	Skolkovo Institute of Science and Technology, Heidelberg University, Novosibirsk State University

OpenDrive数据集

OpenLane Benchmark

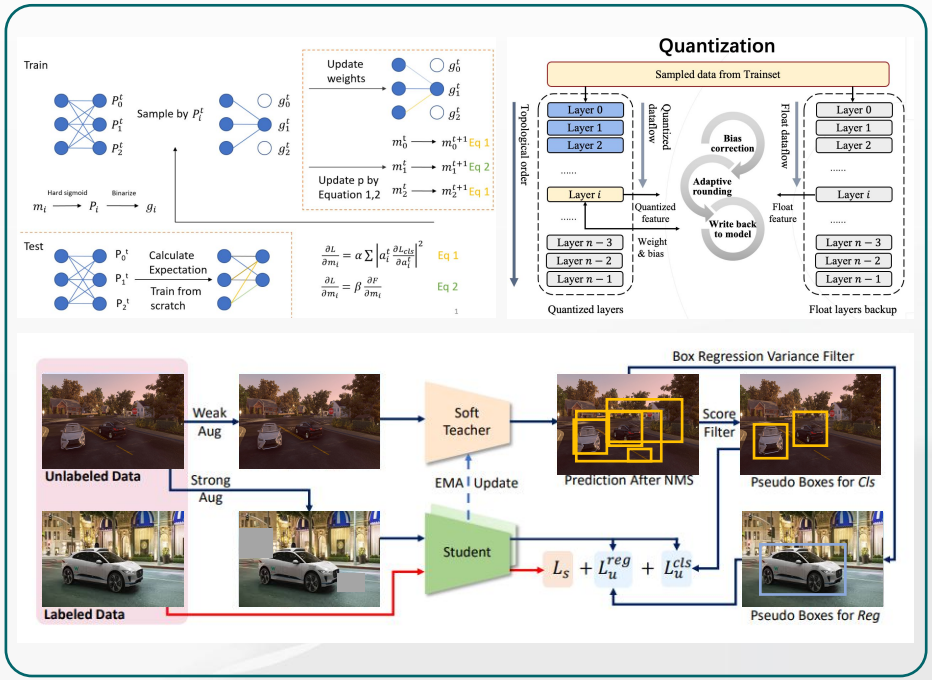





PersFormer工作介绍

算力与平台支撑

数据算法闭环体系 - 核心技术- 模型训练与优化

大模型训练与轻量化部署平台



模型训练

- 通过全景图系统，支持以可视化的方式进行拖拽式建模，实现零代码自动化训练
- 支持自动训练图像分类/物体检测/语义分割模型
- 支持用户手动对训练超参进行调整，并提供模型版本管理功能

模型评测

- 支持单模型评测、多模型对比评测、多数据集对比评测等能力
- 自动可视化显示评测指标

模型轻量化

- 支持大模型轻量化为可部署车端的小模型
- 支持压缩比和精度损失等指标的可视化呈现
- 支持一键打包生成算法SDK

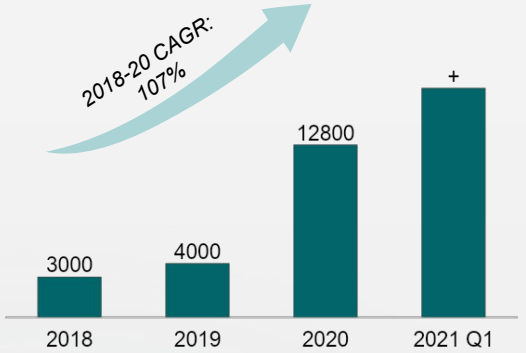
从 模型训练与迭代能力 到 基础平台建设

商汤 sensetime　绝影 SenseAuto

生产单个亿分之一精度算法模型所需的人力、算力 变化

	研究员人数	所需时间	所需算力
2015	10	6个月	100
2020	1	2天	50

累计的算法模型呈指数型增长



数千倍的模型生产效率提升，平台优势显现

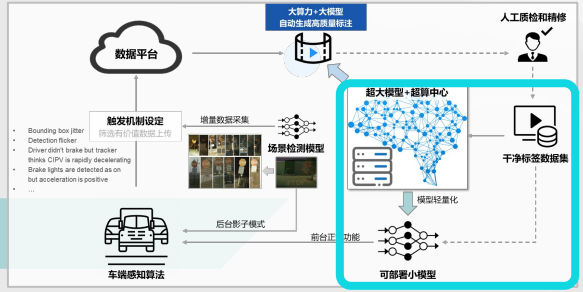
数据平台

AI芯片

AI算力中心

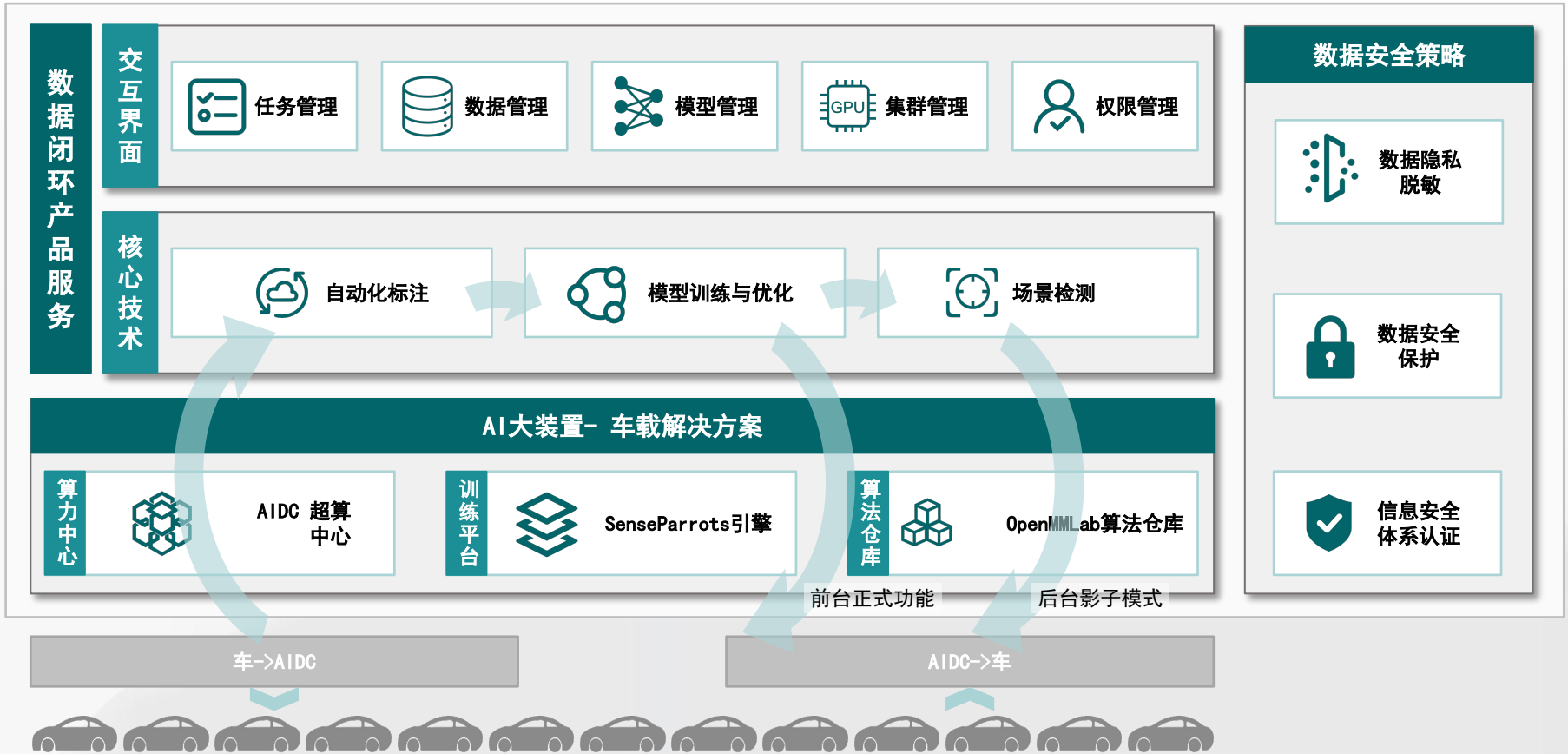
AI训练平台
SenseParrots

AI开源框架
OpenMMlab



商汤AI大装置：人工智能基础设施

SenseAuto Empower - 数据闭环产品架构



AI大装置车载解决方案 | 算力中心- AIDC超算中心（1/2）

商汤AI数据中心赋能智能汽车基础设施

项目规模	建设容量	等效5000个8kW机柜
	项目投资	总投资约56亿元，其中固定资产超过40亿元
技术性能	算力领先	国际领先算力3740Pflops 媲美IDC的存储力160PB
	性能领先	全球视觉最高并行训练效率
应用赋能	10+	头部行业
	100+	科研单位
	1000+	行业企业

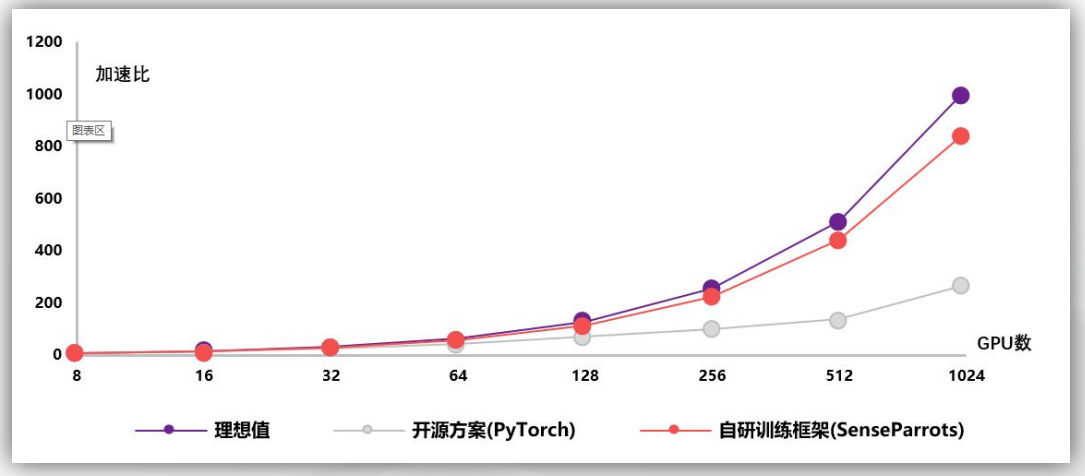
AI大装置车载解决方案 | 算力中心- AIDC超算中心（2/2）

商汤AIDC拥有全球领先的硬件配置

	特斯拉Dojo（美国）	鹏城云脑（深圳）	商汤平台（上海临港）
总算力 **（AI峰值速度）**	1.8EFLOPS	1EFLOPS	**3.74EFLOPs**
存储	10PB	64PB	**160PB**
未来扩容路径	英伟达A100显卡	对华为自研AI芯片、Altas900集群专用属性突出，兼容性较低、扩容难度大	**与主流环境兼容性更高** **与国内芯片厂商深度合作和联合研发**

自建原创训练框架 Parrots

SenseParrots 支持**两千卡**规模的并行训练。在千卡规模上，并行效率超过90%，远超主流框架 PyTorch



千卡并行效率

100%	理想值
91.1%	自研训练框架 (SenseParrots)
29.7%	开源方案 (PyTorch)

60秒训练**AlexNet**

业界最先达到训练最快速度：

在GPU集群上，ImageNet数据集训练速度 1 epoch/s

OpenMMLab

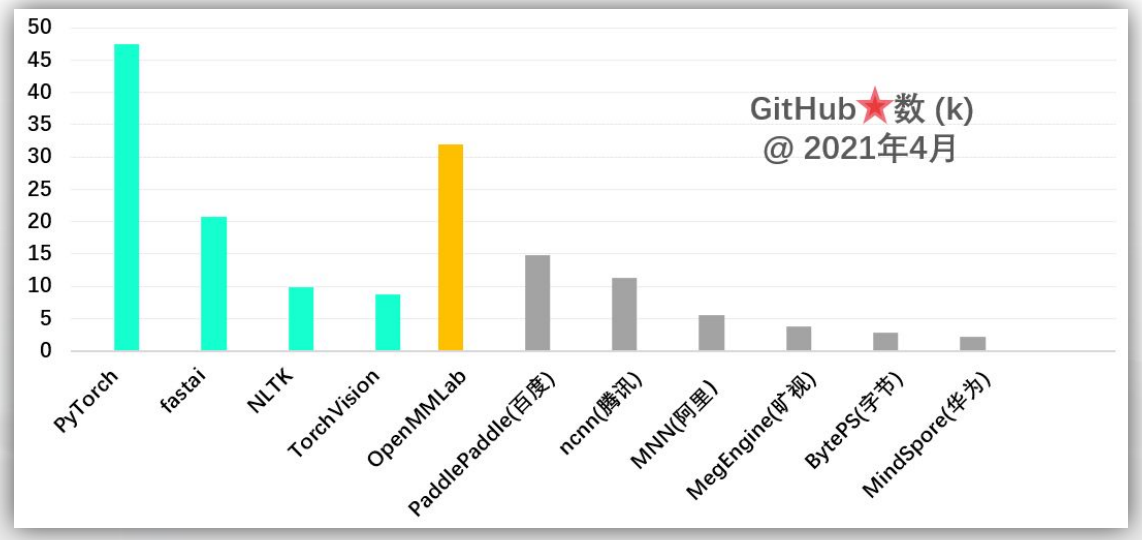
人工智能算法开源体系

收获 **32,000+** GitHub星标

开放超过**140**种算法

1,200个模型

供科研人员和行业应用开发者使用




商汤
sensetime
绝影
SenseAuto
行业标杆分析
MOVE SMART WITH AI

特斯拉 - 数据算法闭环体系 – 解决长尾问题与大模型部署

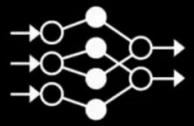
*根据特斯拉 AI Day视频、公开论文数据整理得到

特斯拉

- 60亿标注信息
- 250万视频
- 1.5 PB存储规模
- 1000全职标注人员

- Transformer + RNN + 多任务学习
- 训练时长200+ epoch
- 网络参数量 1000M

- Dojo 超算中心
- D1芯片，算力362 TFLOPs
- 带宽10TBps/dir.
- 功耗400W TDP
- 双SoC, 144 TOPS



量化部署　　车端感知

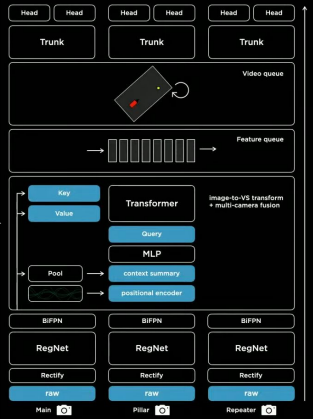
- 感知性能业界最优
- 车端模型推理评估
- 1000次/周，3000 FSD车辆

车队

超大规模数据

- 50+车队，24/7采集
- 100万视频，10亿+标注

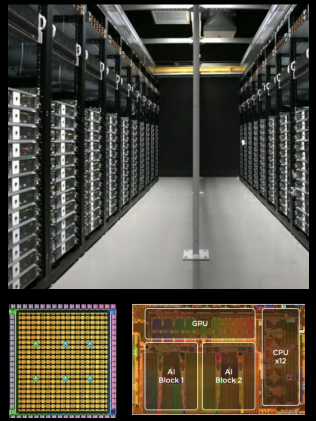
- 公开最大数据集Waymo
- 20万帧数据，5小时
- 1200万标注信息



大模型训练

- Transformer超大规模网络经验
- 训练时长500+ epoch
- 网络参数量 5000+M

- 网络EfficientNet, 结构单一
- 网络参数量100M

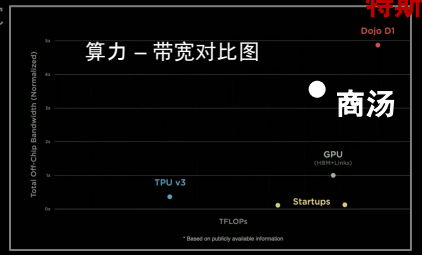


超大规模集群

- AIDC临港超算中心
- 20000+ GPUs
- 数据存储规模160PB
- 总算力 3.74EFLOPs

- 鹏城云脑/华为自研芯片
- 算力1EFLOPS, 存储64PB

带宽



算力

商汤

行业对比

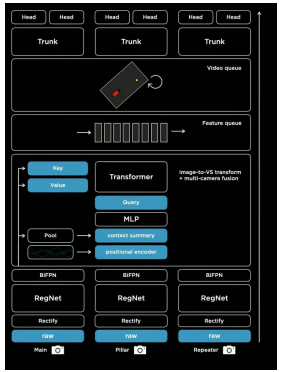
特斯拉借助超大规模网路训练与海量数据，加持卓越的硬件架构，做到业界**绝对领先**。

商汤凭借**SenseCore大装置**，逐步积累实力。

One-page summary

Vision

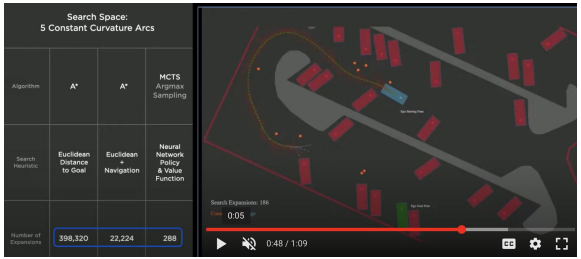
绝对领先



- Multi-task (still new?)
- **Transformer - nV fusion**
  - 解决遮挡和不同视角投影变换
  - smart summon
- **Spatial-temporal**
  - 增加视频前后信息，更好的帮助感知
  - 有助于后续规控
  - c.f. HDMap工作

Planning/Control

- MCTS + policy net节省搜索空间

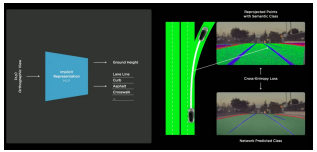


首次提出；Google/Waymo/Uber做的更好

Labelling

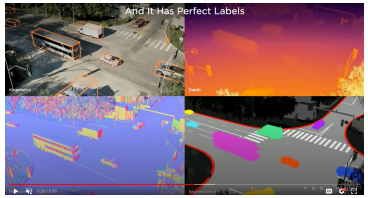
绝对领先/ScaleAI

- 手工标注/第三方标注
  - 早期使用；但效率很低
  - 现在：in-house, **1k 标注员工（非外包）**
- 自动标注
  - scalable
  - **GT真值**：借鉴NeRF思想，self-supervised方式
- **数据规模**
  - 60亿label (含vel/depth)
  - 250w video clip
  - 1.5 PB存储
  - 高质量数据(diverse, clean, large)
  - **Ref: waymo**
    - 20w frame, 5 hours
    - 12M labels



Simulation

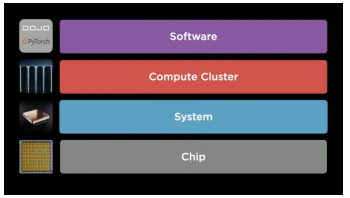
英伟达/谷歌/腾讯



- 仿真提供了大量cornercase数据，帮助算法性能迭代
- 为了仿真和现实场景逼真，做了**五方面**工作；其中场景复原：**neural rendering**

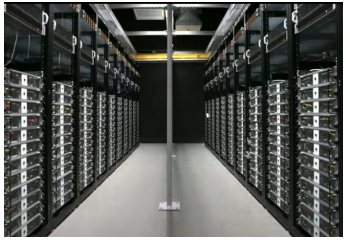
Hardware/Infra/Dojo

感觉很牛逼



- 芯片D1 (DPU)
- 系统
- 计算集群
- 软件架构

各种牛逼数字



- 硬件 3.0 做AI 评估
- 每周运行 100 万次
- 3 个数据中心
- 超过 3000 块 HW3.0 主板组成

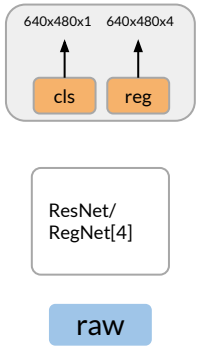
Sum-up: the roadmap in Tesla Vision over the years

- Algorithm
- Product/software/version
- Labelling

2016~

- Regular Network
- 2D detector
- Software 1.0

- manual labeling
- image space labeling



[1] Tesla CVPR 2021 workshop video
[2] Tesla AI day
[3] Smart Summon released on Sep. 26th 2019
[4] Radosavovic, Ilija, et al. "Designing network design spaces." CVPR 2020

Sum-up: the roadmap in Tesla Vision over the years

- Algorithm
- Product/software/version
- Labelling

2016~

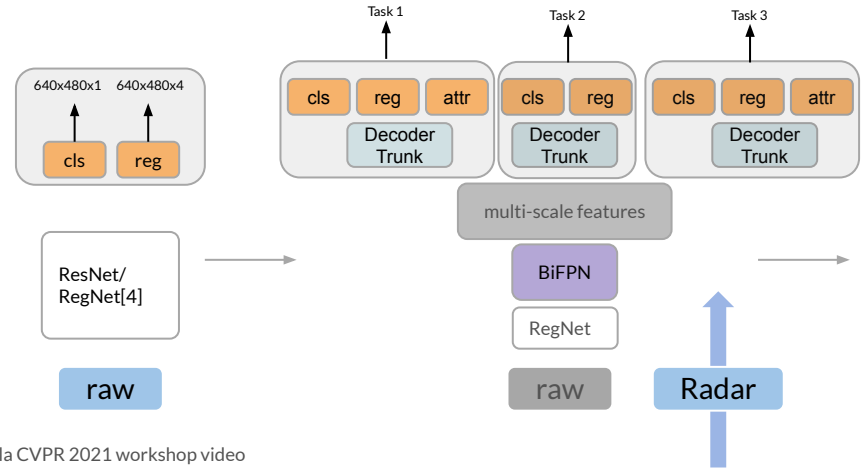
- Regular Network
- 2D detector
- Software 1.0

- manual labeling
- image space labeling

2018-2019

- **Multi-task learning** - "HydraNets"
- Autopilot 4.0

- manual labeling
- vector space labeling



[1] Tesla CVPR 2021 workshop video
[2] Tesla AI day
[3] Smart Summon released on Sep. 26th 2019
[4] Radosavovic, Ilija, et al. "Designing network design spaces." CVPR 2020

Sum-up: the roadmap in Tesla Vision over the years

- Algorithm
- Product/software/version
- Labelling

2016~

- Regular Network
- 2D detector
- Software 1.0

- manual labeling
- image space labeling

2018-2019

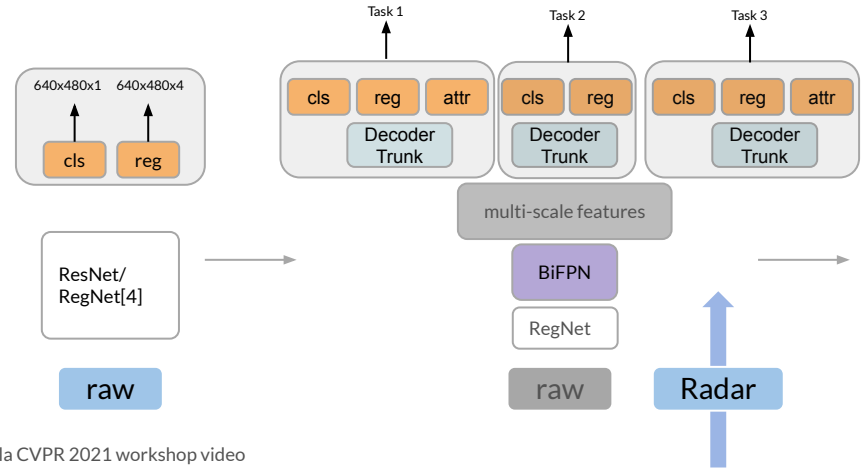
- **Multi-task learning** - "HydraNets"
- Autopilot 4.0

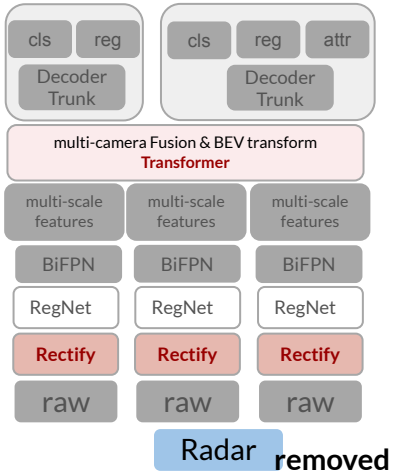
- manual labeling
- vector space labeling

2019-2020

- **Fusion** - Smart Summon [3]
- **Transformer**
- Software 2.0
- Radar removed [1]

- auto labeling
- vector space labeling





[1] Tesla CVPR 2021 workshop video
[2] Tesla AI day
[3] Smart Summon released on Sep. 26th 2019
[4] Radosavovic, Ilija, et al. "Designing network design spaces." CVPR 2020

Sum-up: the roadmap in Tesla Vision over the years

- Algorithm
- Product/software/version
- Labelling

2016~

- Regular Network
- 2D detector
- Software 1.0
- manual labeling
- image space labeling

2018-2019

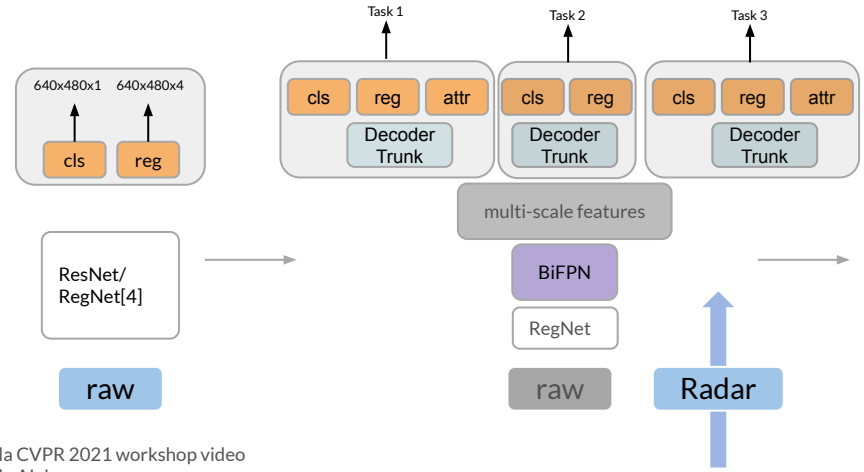
- **Multi-task learning** - "HydraNets"
- Autopilot 4.0
- manual labeling
- vector space labeling

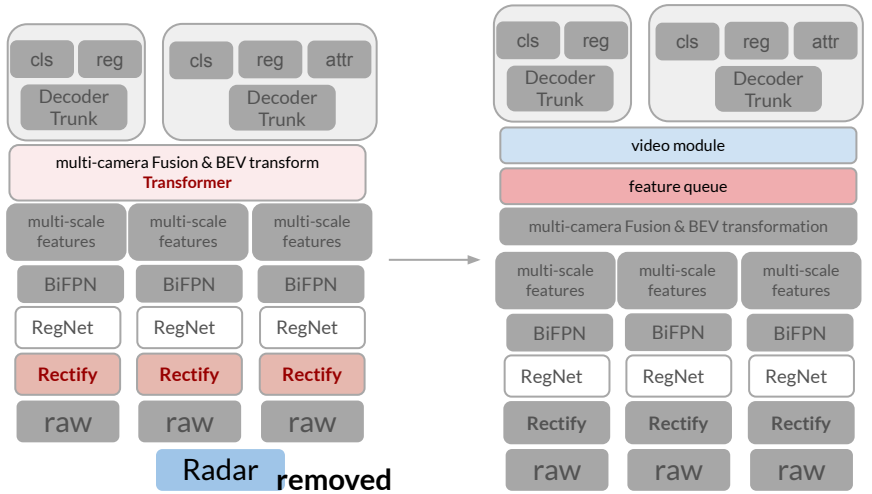
2019-2020

- **Fusion** - Smart Summon [3]
- **Transformer**
- Software 2.0
- Radar removed [1]
- auto labeling
- vector space labeling

2021 [2]

- Spatial-temporal
- Video module
- feature queue
- **FSD beta 9.0/10.0**
- auto labeling
- vector space labeling





[1] Tesla CVPR 2021 workshop video
[2] Tesla AI day
[3] Smart Summon released on Sep. 26th 2019
[4] Radosavovic, Ilija, et al. "Designing network design spaces." CVPR 2020

Take-aways

1. 自动驾驶现状

- 自动驾驶算法大致分为感知、决策两个模块
  - 感知内部各种算法：单目物体检测/车道线检测是基石；nV/fusion融合方案是核心
  - 感知算法性能 简单场景大家做的都差不多；红海市场
  - 决策模块蓝海市场，关键是无法在真 实场景中闭环测试

2. 自动驾驶未来：数据算法闭环体系

- 数据算法闭环体系是未来自动驾驶主流趋势
  - 依靠大模型、基模型提高 feature representation能力；解决泛化能力与 长尾分布
  - Scalable数据量是必要因素；自 动标注技术是体现自动驾驶技术的核心壁垒
  - 算力与深度学 习框架支持是 实现闭环体系的重要保障


商汤 sensetime
绝影 SenseAuto
坚持原创，让 AI 引领人类进步
Email: lihongyang@senseauto.com
Homepage: lihongyang.info
MOVE SMART WITH AI