安信证券 ESSENCE SECURITIES

2023 年 08 月 06 日
汽车零部件 II

行业深度分析

证券研究报告

从特斯拉迭代历程看智能驾驶算法升级趋势

投资评级 领先大市-A
维持评级

首选股票	目标价（元）	评级

行业表现

资料来源：Wind 资讯

升幅%	1M	3M	12M
相对收益	-4.9	20.6	-5.5
绝对收益	-1.7	20.4	-7.4

徐慧雄 分析师
SAC 执业证书编号：S1450520040002
xuhx@essence.com.cn

李泽 分析师
SAC 执业证书编号：S1450523040001
lize@essence.com.cn

相关报告

AI 大模型在自动驾驶中的应用	2023-05-04
汽车轮胎行业系列报告之一：量价齐升叠加成本下降周期，看好中国胎企份额持续向上	2023-04-21
智能网联汽车建设正加速，特定场景商业模式已完成闭环	2023-03-15
智能汽车 2023 年度策略（I）：座舱迈入 2.0 时代，车机域控格局或将再重塑	2022-12-12
线控底盘：实现高阶自动驾驶的必要条件，各环节将迎加速量产期	2022-10-29

供给端特斯拉 FSD Beta 北美渗透率超过 22%，商业化拐点将至。特斯拉 FSD 是在 Autopilot 基础上推出的高阶自动驾驶功能，2020 年 Q3 正式发布 FSD Beta，目前已推送至 V11.4.6 版本。随着软件端不断成熟，FSD 商业化进程提速：**1）测试用户：**截至 2023 年 3 月 FSD Beta 测试人数超过 40 万，在北美渗透率超过 22%。**2）行驶里程：**截至 2023 年 6 月，FSD Beta 累计行驶里已超过 3 亿英里。自 2023 年 4 月开始 FSD Beta 行程里程加速提升，仅 Q2 单季度提升约 1 亿英里。**3）软件付费：** 2019 年以来经过多轮涨价，目前 FSD 买断价格高达 1.5 万美元。马斯克在推特上预告 FSD V12 将不再是 Beta 版本，并会在北美向所有用户免费试用一个月，我们认为届时 FSD 订阅率有望实现跃升，智能驾驶付费彻底跑通。

算法：从规则主导到神经网络主导，BEV+Transformer 确立行业通用感知范式，端到端大模型有望再次引领行业。2016 年之后特斯拉开始尝试自研算法，并在 2017 年 6 月 Andrej 加入特斯拉之后确立从传统视觉向神经网络模型、数据驱动方向发展的技术路线，逐步构建了 BEV+Transformer 感知范式。1）2018-2019 年特斯拉在停车场智能召唤场景下开始将不同视角下的车道线预测结果投射到 BEV 视角下拼接，采用基于规则的方式进行后融合；2）2020-2021 年特斯拉在 BEV 空间内做特征级融合取代基于规则的后融合，确立 BEV+Transformer 感知范式，自动驾驶进入大模型时代；3）2022 年从 BEV 升级至 Occupancy（相当于 4D BEV，增加高度信息），进一步提升对一般障碍物的识别能力。根据马斯克推特，FSD V12 将是“端到端”的自动驾驶模型，在规划决策端应用神经网络，将大幅提升自动驾驶车辆与交通参与者的交互能力。

数据：数据闭环+超算中心造就 FSD 极致的迭代速度，是特斯拉在自动驾驶上的核心竞争力。复盘特斯拉对数据闭环体系的构建可以分为以下两个阶段：**1）2016-2019 年：**特斯拉早在 2016 年首创影子模式，组建千人的标注团队，数据引擎初步构建。采购英伟达 GPU 自建数据中心，规模尚小，模型单次训练所需时间较长。2）2020 年至今：特斯拉升级版数据引擎在标注、仿真、云端计算资源三个方面大幅升级，数据闭环系统趋于完善。随着感知模型向 BEV 迭代，特斯拉数据标注也从 2D 人工标注发展至 4D 自动标注，大幅提高标注效率。2019 年以来，特斯拉基于英伟达 GPU 部署的数据中心算力持续提升，并同步自研 Dojo 超算中心。2023 年 7 月 Dojo 超算中心正式投产，FSD 迭代速度有望进一步大幅提升。

硬件：HW4.0 版本发布，芯片性能、传感器配置全面升级。特斯拉自 2019 年发布 HW3.0 起搭载自研的 FSD 芯片，2023 年 HW4.0 已开始量产，在芯片性能及传感器配置方面全面升级。**1）传感器配置：**摄像头精度提升并新增 4D 毫米波雷达；**2）算力：**自动驾驶大模型推动车端算力需求的提升，预计 HW4.0 总算力由 144Tops 提升至 400-500Tops；**3）内存：**Transformer 大模型对内存消耗大幅提升，为了解决“内存墙”问题，特斯拉从 HW3.0 的 8 颗 LPDDR4 升级至 16 颗 GDDR6，内存容量从 16GB 提升至 32GB，最大内存带宽从 68GB/s 大幅提升至 224GB/s。

相关标的：小鹏汽车，理想汽车，德赛西威

风险提示：技术进步不及预期，商业化进展不及预期等。

内容目录

图表目录

1. 特斯拉 FSD 商业化拐点将至，智驾付费模式有望彻底跑通

特斯拉 FSD（Full Self-Driving）是在 Autopilot 的基础上，推出的高阶自动驾驶功能，是特斯拉树立“高端智能化”品牌标签的重要渠道，目前已迭代至 V11.4.6。特斯拉于 2020 年 Q3 正式发布 FSD Beta（测试版）版本，随后在 2021 年 7 月特斯拉通过重构后的底层算法，采用纯视觉技术路线初步实现了城市 NOA，并针对不良天气影响、无保护左转等 Corner case 进行不断的升级优化。从 2023 年 4 月发布的 FSD Beta 11.3 版本开始，特斯拉统一了城市 NOA 与高速 NOA 的系统架构。**根据马斯克在推特上的多次预告，FSD V12 将是一次具有历史意义的重要更新，同时称 FSD V12 将不再是 Beta 版本。**

表1：特斯拉 FSD Beta 更新历程

时间	版本	特斯拉 FSD 重要 OTA 内容
2021/7/10	FSD Beta v9.0	重构底层算法，改为纯视觉方案并首次新增城市场景
2021/11/7	FSD Beta v10.4	升级道路中紧急车辆检测网络
2021/11/22	FSD Beta v10.5	在影子模式下启用“紧急碰撞避免控制”功能。
2021/12/19	FSD Beta v10.7	改进了物体属性网络，将错误并道减速减少了 50%，误差减少了 19%。
2022/4/13	FSD Beta v10.11	1、显著降低了由于柏油路裂缝、刹车痕迹、雨天等因素造成的误报。 2、将车道拓扑结构关系建模从密集栅格（点的集合）升级为自回归解码器，使用 transformer 神经网络直接预测并逐点连接向量空间中的车道。从而实现预测变道行为，降低了运算量并减少了的错误率。
2022/5/15	FSD Beta v10.12	1、升级了“无保护左转”场景的决策框架，通过增加更多的特征刻画 go/no-go 决策来更好地建模他车对自车决策行为的反应。保证噪点环境下决策结果仍在安全边界内的鲁棒性。 2、在规划车道选择时更好地融合了目标未来的轨迹预测信息，从而降低不舒适转弯行为的概率。
2022/9/11	FSD Beta v10.69	1、在矢量车道神经网络中增加了一个新的 "深度车道引导 "模块。 2、添加了对 Occupancy Network 中任意低速移动体积的控制，从而更精细地控制一些无法用长方体单元表示的物体形状。
2022/12/15	FSD Beta v10.69.3	1、将物体检测网络升级为光子视频流（不做滤波处理），并使用最新的自动标记数据集对所有参数进行了重新训练（特别侧重于低能见度场景）。 2、将 VRU 速度网络转换为两阶段网络。 3、重新制定了自回归矢量车道语法并为矢量车道神经网络增加了一个新的 "道路标记 "模块，使十字路口的车道拓扑结构误差改善了 38.9%。
2023/4/8	FSD Beta v11.3	在高速公路上启用了 FSD 测试版。统一了公路和非公路上的视觉和规划堆栈。
2023/4/17	FSD Beta v11.4	相比于 V11.3 在转弯和变道的表现上有明显的提升（包括当转弯车道被停放的车辆挡住、避免转入公交车道、为避免堵塞提前变道等）。
2023/5/26	FSD Beta v11.4.2	提高对远距离道路特征的检测，对其他车道上的车变道到自己车道的预测成功率提升。
2023/6/14	FSD Beta v11.4.3	1、通过改进车道、道路边缘等的拓扑，进而改进对转弯的控制和平滑度。 2、将车道信息引入占用网络模型，提升对远距离道路特征的识别。 3、提升 cut-in 和 cut-out 的准确率。
2023/6/20	FSD Beta V11.4.4	1、改进对目标车道中的车辆的建模，改进了紧急情况下更换车道的表现。 2、改进对对面来车轨迹的预测，进而改进狭窄的 free space 场景下，对迎面而来的车辆的处理。 3、利用更多的运动学数据来改进对人行横道附近 VRU 未来轨迹的预测。 4、将新物体真值自动标注系统拓展到 Non VRU，进而提高对远距离异形车辆的识别。
2023/7/22	FSD Beta V11.4.6	改进了自动紧急制动系统（AEB），除了车辆和行人之外，其他被占用网络检测到的物体现在也会激活自动制动系统。

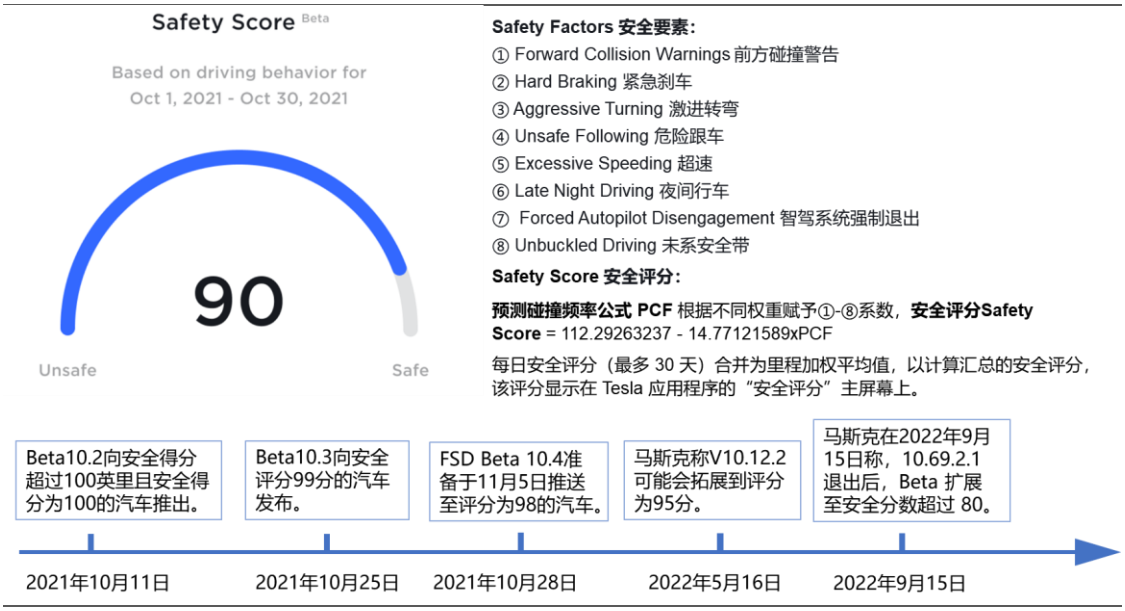
资料来源：特斯拉官网，安信证券研究中心

FSD Beta 在北美测试用户已超 40 万，行驶里程加速提升，我们认为特斯拉 FSD 商业化拐点将至，智能驾驶付费模式有望彻底跑通。

测试用户：马斯克早在 2015 年首次官宣特斯拉将推出 FSD 完全自动驾驶，2016 年 Q3 在官网上线 FSD 选装包，彼时尚无具体功能说明。直到 2020 年 10 月 21 日，特斯拉正式发布 FSD Beta 测试版本，但仅向北美 Early Access 早鸟用户推送。在 2021 年初举行的财报电

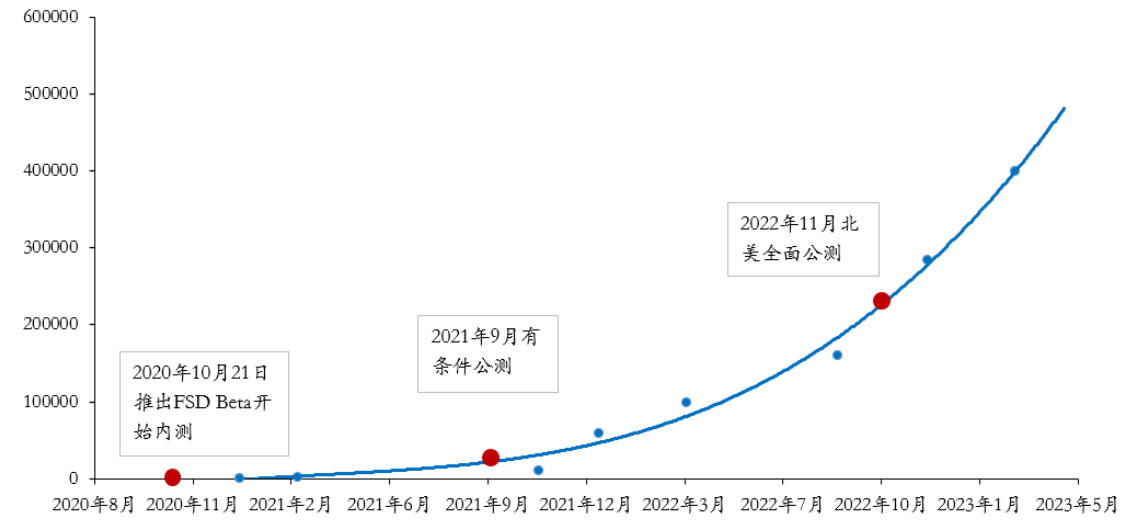
话会议上，马斯克表示，截至 2021 年 1 月，已经有近 1,000 名车主在公共道路上参与 FSD 测试，至 2021 年 3 月，这一数字已提升至 2000 名。在数千名早鸟用户历经一年的内测，伴随着 2021 年 7 月发布重要版本 FSD Beta V9，在 2021 年 9 月 FSD Beta 在北美开始进行有条件的公测，但仅安全评分达到 100 的车主才可获得测试资格。两个月后，2021 年 11 月 FSD Beta 测试者的数量大幅提升至 1.17 万，随后 2022 年 1 月/4 月/9 月测试人数分别达到 6 万/10 万/16 万。随着 FSD Beta 版本持续迭代、系统可靠性不断提升，特斯拉对于获得 FSD Beta 测试资格的安全评分标准不断放松，根据马斯克的推特，2022 年 9 月对安全评分的要求已放宽至 80 分。至 2022 年 11 月 24 日，特斯拉向北美地区所有购买 FSD 用户推送 FSD Beta 测试功能，标志着 FSD Beta 在北美进入全面公测，参与测试人数随之大幅提升，截至 2022 年 12 月末测试人数达到 28.5 万。**根据特斯拉官方推特，截至 2023 年 3 月 2 日 FSD Beta 测试人数超过 40 万，根据 Marklines 数据，截至 2023 年 2 月底，北美特斯拉保有量约为 185 万，对应渗透率达到 22%。**（FSD 软件需要在 HW3.0 平台上才可以启动，特斯拉 2019Q2 之后生产的车辆才搭载 HW3.0，但特斯拉可以为老车主将硬件免费升级至 3.0 平台，因此此处渗透率计算按特斯拉在北美全部的保有量计算）

图1. 获得 FSD Beta 测试资格的安全评分要求不断降低



资料来源：Tesla 官网，马斯克推特，安信证券研究中心

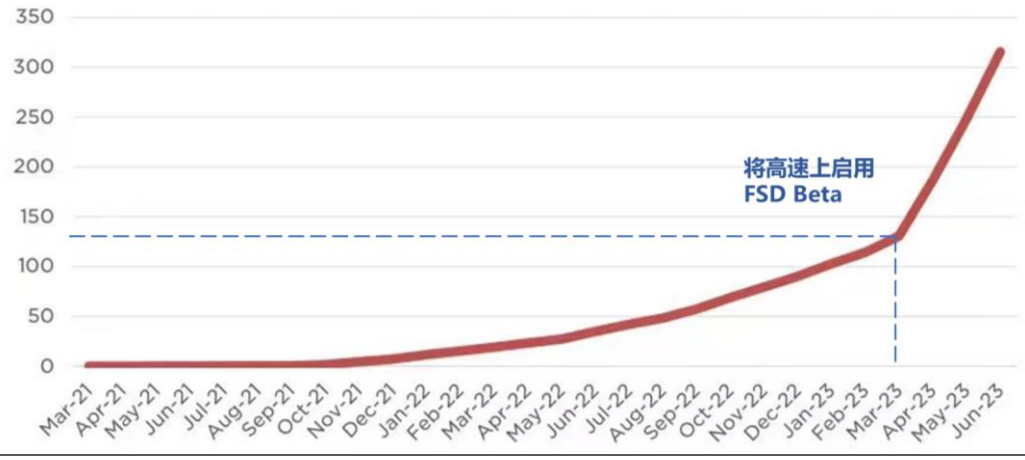
图2. FSD Beta 测试用户数已超过 40 万



资料来源：Tesla、马斯克推特，安信证券研究中心

行驶里程：根据特斯拉 2023 年二季度业绩说明会，截至 2023 年 6 月，FSD Beta 累计行驶里程已超过 3 亿英里。其中，自 2023 年 4 月开始 FSD Beta 累计行程里程加速提升，仅 Q2 单季度提升约 1 亿英里，主要系 FSD 订阅量的上升及从 2023 年 4 月开始的 V11.3 在高速上启用了 FSD Beta。需要注意的是，在 2023 年 4 月之前，高速场景并未统一到 FSD Beta 技术栈中。

图3. FSD Beta 行驶里程超过 3 亿英里



资料来源：特斯拉 2023Q2 业绩说明会，安信证券研究中心

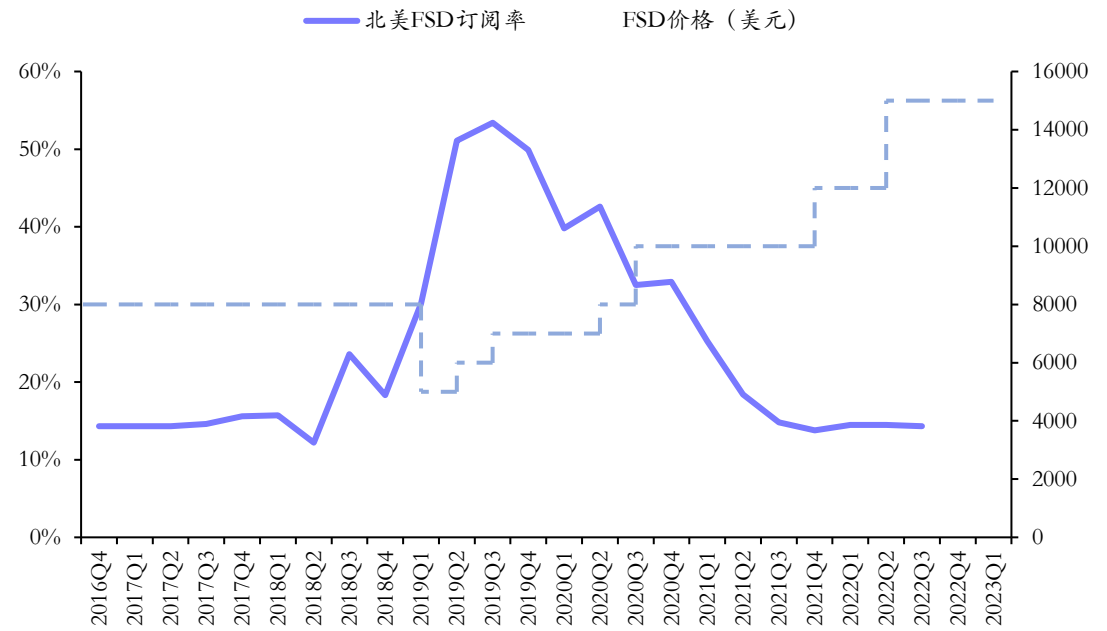
软件付费：特斯拉 FSD 具有“期货”属性，自 2016 年发布以来已经过多轮价格调整，2019 年 4 月激活 FSD 功能仅需要一次性支付 5,000 美元，而目前 FSD 买断价格已上涨至 1.5 万美元。同时自 2021 年 5 月起，FSD 同时支持订阅的方式进行购买，基础 AP 用户订阅价格为 99 美元/月，已购买加强 AP 的用户订阅 FSD 价格为 199 美元/月。根据 TroyTeslike 调研数据，2019 年以来随着 FSD 购买价格逐步上涨及 Model 3/Y 中低端车型成为销售主力，FSD 在北美的单季度渗透率有所下滑。随着 FSD Beta 功能体验逐步完善，2023 年下半年以来特斯拉通过对 FSD 进行有条件优惠等方式扩大用户基数。2023 年 7 月 7 日特斯拉升级引荐计划，如果用户通过推荐购买 Model 3/Y，可以免费试用三个月 FSD Beta；如果通过推荐购买 Model S/X，可以免费试用六个月。**同时，马斯克在推特上表示当 FSD 达到足够流畅时会在北美向所有用户免费试用一个月，我们认为这或会在 FSD V12 发布后实现，届时 FSD 订阅率有望实现跃升，智能驾驶付费彻底跑通。**

表2：特斯拉 FSD 自 2019 年 4 月以来价格持续上涨

收费方式	时间：	方式描述	价格（均为购车时就加装的价格）		
完全买断制	2019 年 4 月前	购车时可以选择加强 AP，获得 FSD 功能必须在购买加强 AP 的基础上额外付费	加强 AP 价格：	$5,000	
			FSD 价格：	$8,000（含必须的加强 AP 价格和 FSD 额外付费$3,000）	
	2019 年 4 月至 2021 年 5 月	基础 AP 标配，不提供加强 AP，FSD 可单独选装	日期	FSD package 买断价格	
			2019 年 4 月	$5,000	
			2019 年 5 月	$6,000	
			2019 年 8 月	$7,000	
			2020 年 7 月	$8,000	
			2020 年 10 月	$10,000	
买断+订阅双机制	2021 年 5 月至今	客户可以选择直接购买 package，也可以选择按月订购，订购价格取决于客户的 AP 类型	加强 AP 价格：	在 2020 年 9 月提供限时选项，$5,000 2022 年 6 月起正式成为可选项，$6,000	
			日期	FSD package 买断价格	FSD 订阅价格
			2022 年 1 月之前	$10,000	加强 AP 用户：$99/月 基础 AP 用户：$199/月
			2022 年 1 月	$12,000	
			2022 年 9 月	$15,000	

资料来源：特斯拉官网等，安信证券研究中心

图4. 北美 FSD 单季度订阅率随着受到价格上涨影响有所下降



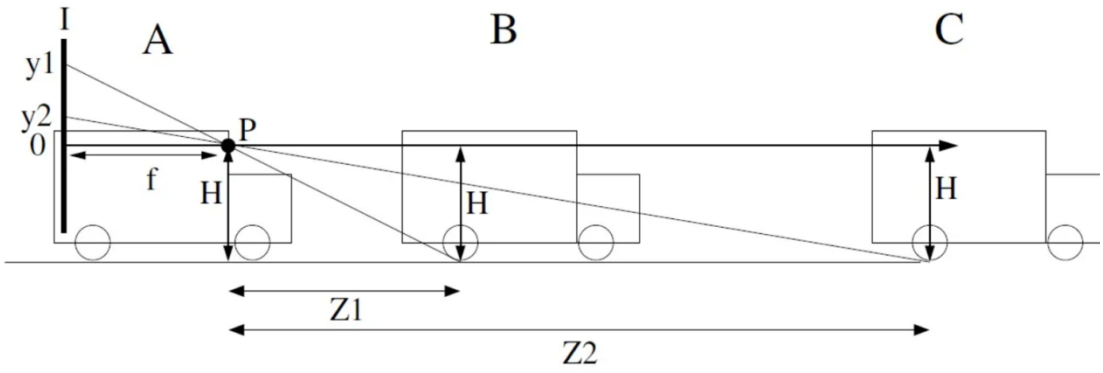
资料来源：TroyTeslike，notateslaapp，安信证券研究中心

2. 算法：BEV+Transformer 确立行业通用感知范式，端到端大模型有望再次引领行业

2.1. 2018 年之前：从与 Mobileye 合作到初步尝试自研

2014-2016 年间特斯拉与 Mobileye 深度合作，由 Mobileye 提供感知算法，主要基于传统机器视觉技术，依靠大量人工手写规则。针对每一类 ADAS 任务，Mobileye 都设计了复杂的机器视觉算法，并且在工程层面进行长期的优化，结合专用芯片，最终达到效率和可靠性的平衡。以 Mobileye 的经典测距算法为例，它使用前方车辆的车轮和地面接触点作为检测点，在假定地面水平的情况下，利用镜头的焦距 f、相机离地距离 H、成像高度 y 等易于测量的数据，可以估算出车辆距离本车的距离。**在这一阶段，特斯拉基于 Mobileye 方案的 AP1.0 系统陆续实现了车道偏离预警、主动巡航控制、自动变道、自动泊车等功能。**

图5. Mobileye 经典的测距方法，利用几何关系计算 A 车与 B，C 车的距离



资料来源：Vision-based ACC with a Single Camera: Bounds on Range and:Range Rate Accuracy，安信证券研究中心

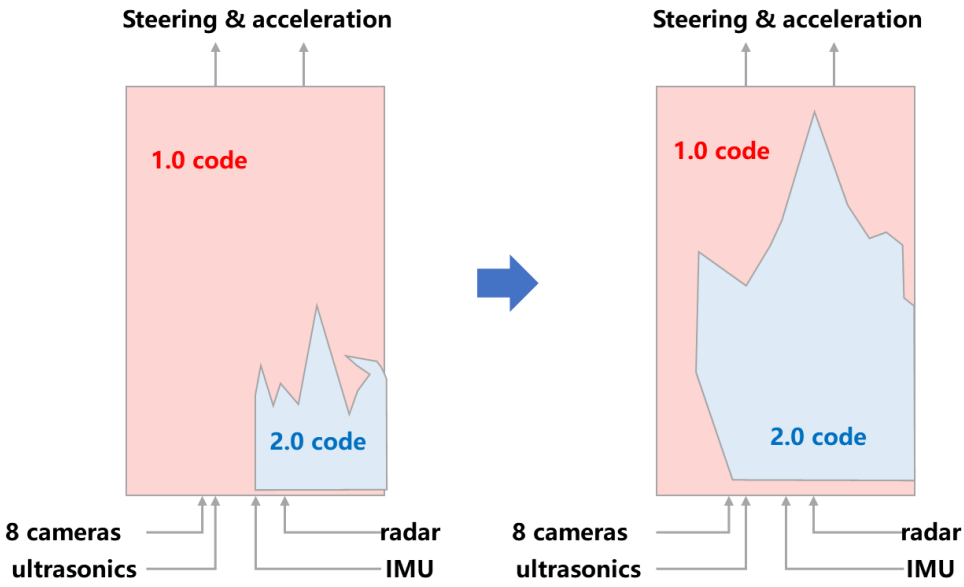
2016 年-2017 年特斯拉开始逐步探索自研自动驾驶算法。2016 年特斯拉和 Mobileye 合作关系破裂后在硬件端转向英伟达，同时自研软件算法。2016 年 10 月 HW2.0 量产，而软件层尚未推出，直到 2016 年 12 月 31 日特斯拉发布 Autopilot8.0 版本，辅助驾驶功能才重新上

线，但相比于 AP1.0 系统功能上出现了明显的回退，至 2017 年 3 月推送的 8.1 版本 AP2.0 系统基本达到了 AP1.0 系统的功能体验。同时，2016-2018 年间特斯拉自动驾驶团队构成也发生了多次变化，2016 年 12 月，特斯拉 Autopilot 原总监 Sterling Anderson 离职，苹果 Swift 语言之父 Chris Lattner 接任，带领 AP2.0 的研发，但仅半年后 Lattner 宣布离职。在这一时期特斯拉几乎不对外披露软件算法技术进展，**但值得注意的是，2016 年底开始特斯拉的 vision 小组与机器学习小组也开始在技术上为 Autopilot 的开发提供支持，说明特斯拉已经开始尝试将 AI 引入自动驾驶的应用中。**

2.2. 2018 年之后：从后融合到特征级融合，大模型赋能下引领行业

2017 年 6 月 Andrej 加入特斯拉后，主导特斯拉自动驾驶算法从基于传统视觉（规则的方式）向神经网络模型、数据驱动的方向发展。Andrej 将传统视觉称之为 Software1.0，指实现某一个功能依靠既定代码逻辑，可以理解为给定目标，程序员设定好一条固定达到目标的路径。以数据驱动、依靠神经网络的模型被称为 Software2.0，给定目标结果，程序员设定网络框架，通过计算资源搜索程序空间的子集（给定目标值，利用反向传播和梯度下降实现），进而找到这条具体的、最高效的路径。**特斯拉自动驾驶算法进化过程是 2.0 软件逐步“吞噬”1.0 软件的过程，从一开始规则主导，部分神经网络辅助；到二者交叉，部分模块神经网络、部分规则，再到神经网络完全主导，用“one model”统一全栈。**

图6. 特斯拉自动驾驶模型逐步由规则主导到神经网络主导

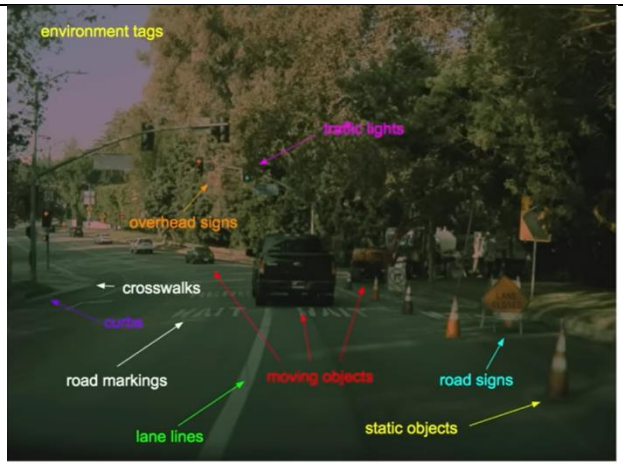


资料来源：Building the Software 2.0 Stack (Andrej Karpathy)，安信证券研究中心

2.2.1. 2018-2019：使用多任务网络提高模型效率，在 BEV 空间下进行后融合

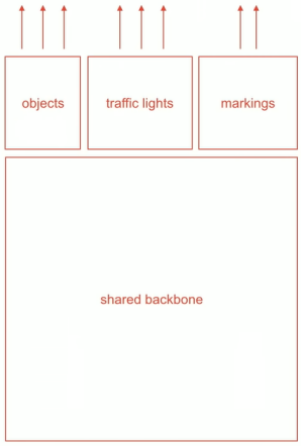
构建 Hytranets 多任务网络提高自动驾驶感知模型效率。在 2018-2019 年期间，行业中应用神经网络完成自动驾驶感知任务的方式是针对单个任务进行网络设计，即一个神经网络结构只对应一个感知任务的实现。自动驾驶中同时存在非常多感知任务（尤其从高速进入城市场景，环境复杂度大幅提升），如果为每一个任务单独设计一个神经网络极其耗费资源。特斯拉的解决方案是设计一个 Hydranets 多任务网络，有一个共享的 backbone 骨干网络，再输出多个任务。这样设计最核心的好处在于节约计算资源，一方面在训练端，针对单个任务进行微调时不需要对共享网络进行重新训练；另一方面在车端进行推理时不同任务共享特征提取结果从而避免重复计算。

图7. 自动驾驶感知同时存在多种任务



资料来源： PyTorch at Tesla - Andrej Karpathy，安信证券研究中心

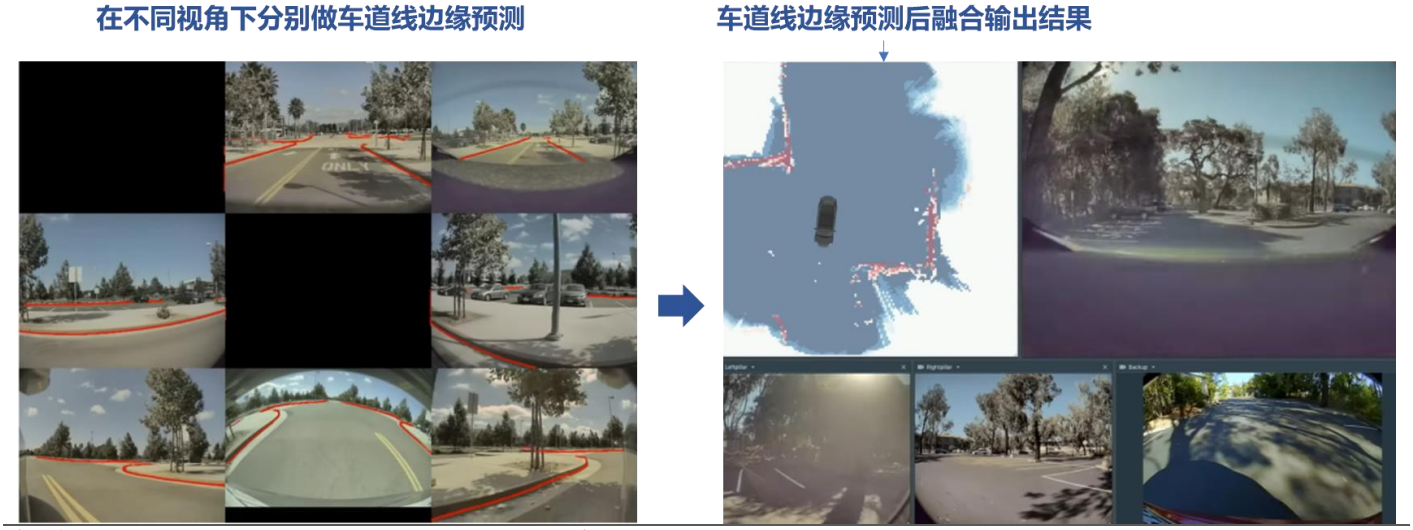
图8. 特斯拉 Hydranets 多任务网络



资料来源：PyTorch at Tesla - Andrej Karpathy，安信证券研究中心

在泊车场景下开始应用 BEV，采用后融合策略对不同视角进行拼接。2019 年 10 月，特斯拉推出停车场智能召唤功能，可控制车辆离开车位、绕过弯角、进行必要的避障到达所选位置。为实现这一功能，车辆需要找到停车场中的可行驶区域，避免碰到道路边缘。特斯拉在不同视角之下完成了车道线边缘的预测，但车辆无法在 2D 透视图中完成后续的规划决策，因此需要将 8 个不同视角下的预测结果“投射”到 BEV 视角下（此时尚未正式提出 BEV 的概念，称其为 Top-down 自上而下的视角）进行拼接，**需要特别注意的是，这个拼接过程是用基于数学规则的方式而非神经网络的方式完成的。**

图9. 在 BEV 空间内进行后融合

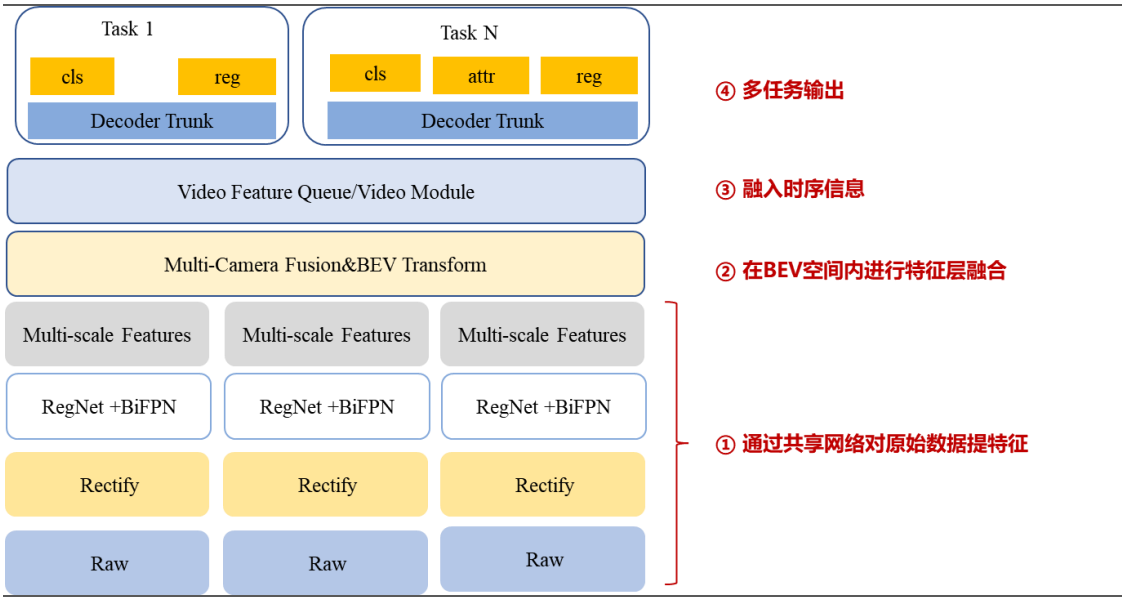


资料来源： PyTorch at Tesla - Andrej Karpathy，安信证券研究中心

2.2.2. 2020-2021：特征级融合取代后融合，BEV+Transformer 架构下，进入自动驾驶大模型时代

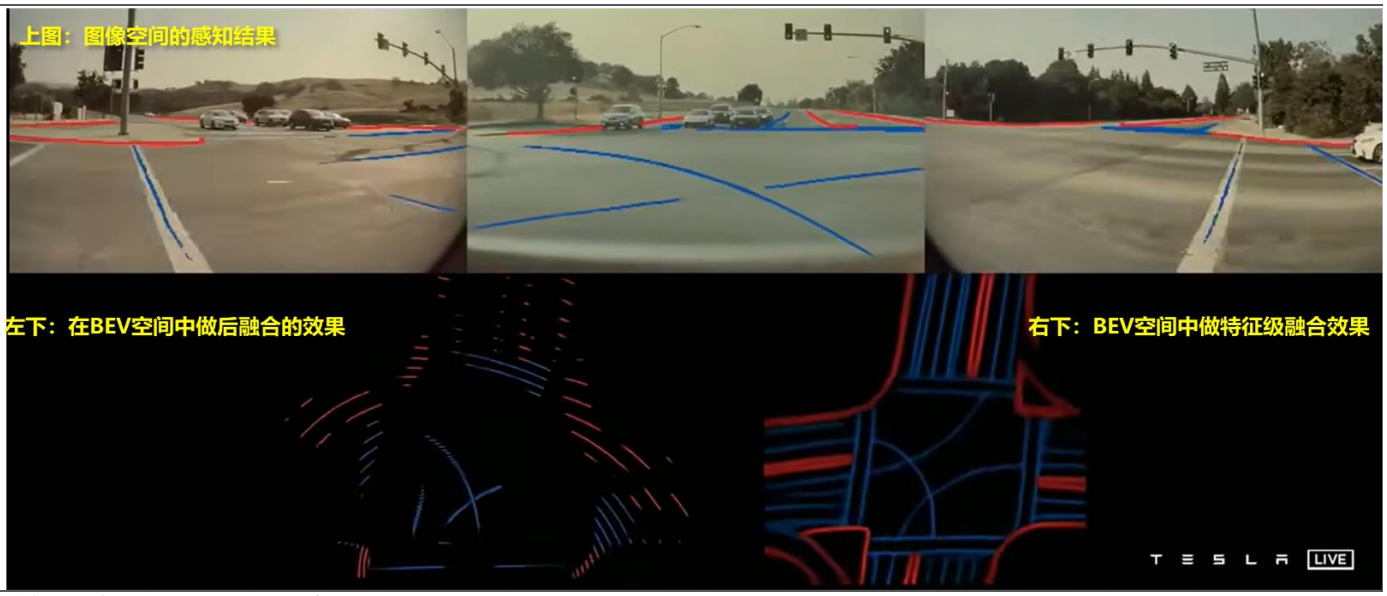
特斯拉利用基于神经网络的特征级融合取代基于规则的后融合，大幅提升感知效果。2020 年特斯拉开始研发 FSD 完全自动驾驶，当自动驾驶从简单的泊车场景向普通城市道路拓展时，后融合的感知结果难以满足要求。一方面基于规则的后融合具有严苛的假设，如地面是完美水平、相机和地面之间不存在相对运动，因此任何的车辆颠簸或者道路有高度变化都会打破这一假设，使得 BEV 输出的图像面临失真。同时，由于透视投影，在 2D 图像中完成不错的感知结果投影到 BEV 空间中精度很差，若要保证远距离区域的精度，就必须要对每一个像素的深度预测非常准确，而这是难以实现的。**为解决这些问题，特斯拉希望能直接利用神经网络输出 BEV 感知结果，自动驾驶感知融合从后融合走向特征级融合。**具体模型框架如下：1）通过 Backbone 共享骨干网络进行特征提取；2）将不同视角下的 2D 特征图通过神经网络转换至 BEV 空间内融合；3）融入时序信息；4）多任务的输出。

图10. 特斯拉 BEV 感知模型架构



资料来源：特斯拉 AI Day，安信证券研究中心

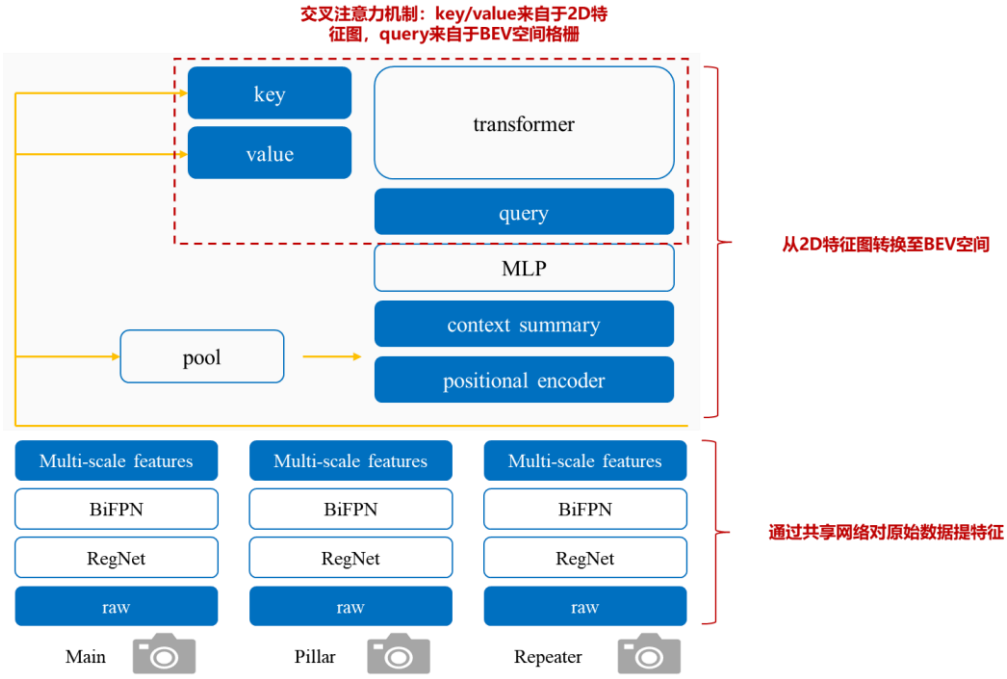
图11. 在 BEV 空间内做特征级融合的效果远好于基于规则的方法在 BEV 空间内后融合



资料来源：特斯拉 AI Day，安信证券研究中心

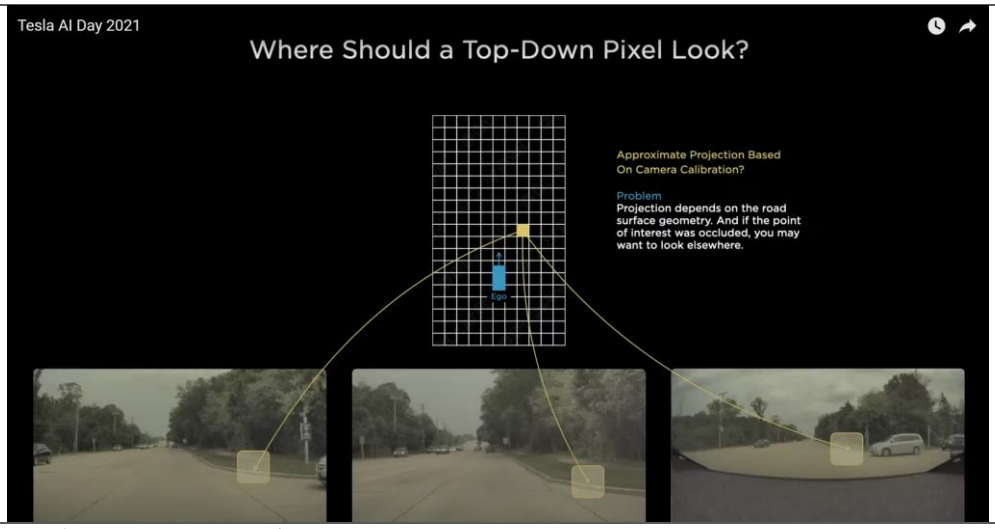
Transformer 交叉注意力机制对于 BEV 空间转换任务适配性较高。利用 Transformer 进行 BEV 空间转换的方法没并有显示的深度估计，而是用注意力机制直接进行不同序列（指 2D 特征图和 BEV 视图）之间的转换。Transformer 的交叉注意力机制中的 Query 和 Key/Value 来源不同，因此天然适配于不同域之间的数据转换。在 2D 特征图向 BEV 空间转换的过程中，首先将 BEV 空间分割成 2D 格栅，之后将它们编码成一组 Query 向量，去不同视角的 2D 特征图中查询对应的点，从而实现空间的转换。根据 2021 年特斯拉 AI Day，通过 Transformer 交叉注意力机制在 BEV 空间内做特征级融合的效果远好于基于规则的方法在 BEV 空间内后融合。

图12. 特斯拉基于 Transformer 的 BEV 空间转换过程



资料来源：特斯拉 AI Day，安信证券研究中心

图13. BEV 转换过程

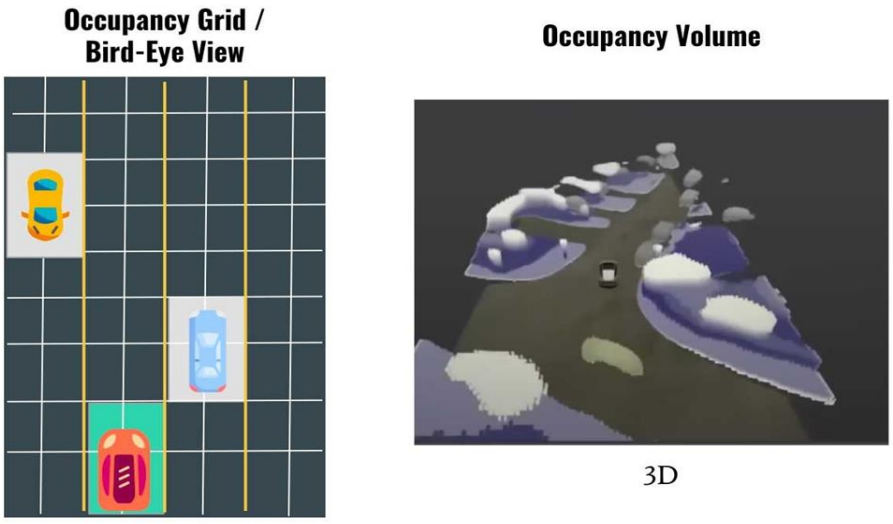


资料来源：特斯拉 AI Day，安信证券研究中心

2.2.3. 2022：升级至 Occupancy 解决一般障碍物识别问题，Lanes Network 进一步完善地图模型

从 BEV 升级到占用网络，进一步提升泛化能力。特斯拉在 2022 年 AI Day 中展现了 Occupancy Network 感知技术。基本的思想是将三维空间划分成体素 voxel（可以理解为微小立方体），再去预测每个 voxel 是被占用还是空闲，通过 0/1 赋值对 voxel 进行二分类：有物体的 voxel 赋值为 1，表示 voxel 被物体占据；没有物体的 voxel 被赋值为 0。实际中的赋值可以是概率值，表示 voxel 存在物体的概率。

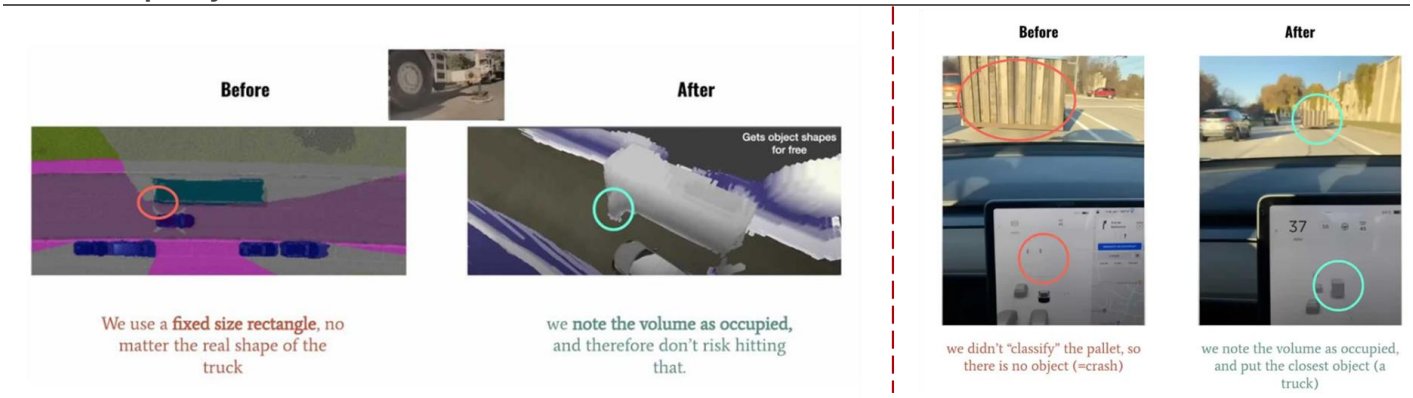
图14. BEV 与占用网络输出结果对比



资料来源：Think Autonomous，安信证券研究中心

占用网络感知技术本质上是为了解决更多的长尾问题。纯视觉方案被质疑的一大问题在于对于没有在训练集中出现过的物体，视觉系统则无法识别，比如侧翻的白色大卡车，垃圾桶出现的路中，传统视觉算法无法检测到。占用网络模型的基本思想是“不考虑这个物体到底是什么，只考虑体素是否被占用”，则从根本上避免了这一问题，大幅提升了模型的泛化能力。

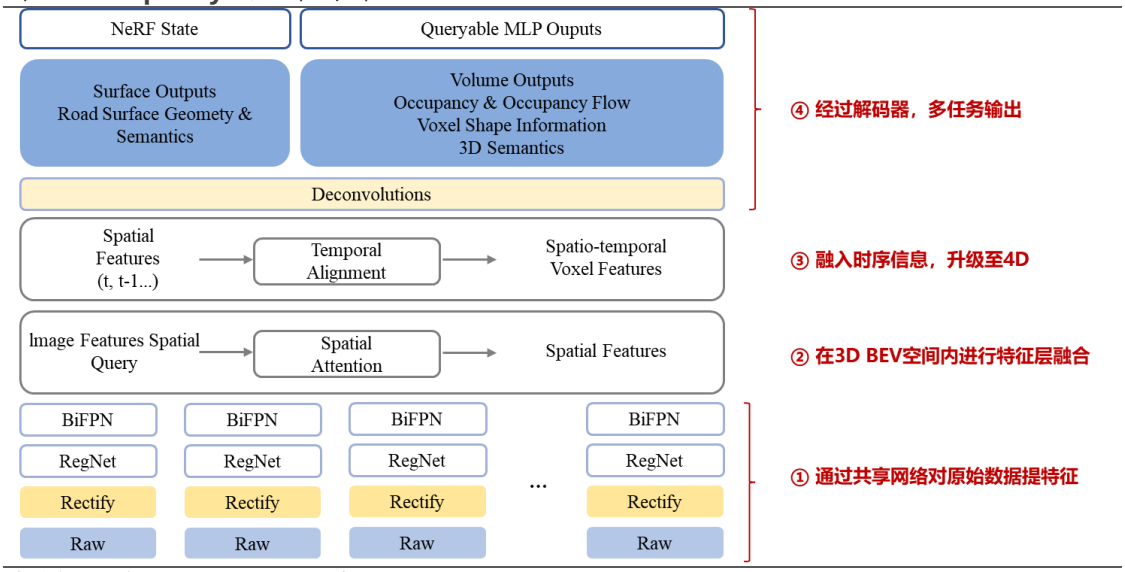
图15. Occupancy 有效解决一般障碍物识别问题



资料来源：Think Autonomous，安信证券研究中心

Occupancy 网络结构与特斯拉 2021 年 AI Day 展示的 BEV 网络结构差异不大，均包括特征提取、利用神经网络进行特征级融合、融入时序信息、多任务的输出四个步骤，事实上 Occupancy 可以看作是 4D 的 BEV。从网络结构上看差异主要体现在：1）Occupancy 模型中进行空间转换时的 Query 是 3D 格栅，BEV 模型中是 2D；2）Occupancy 模型可以直接解码出网格的占用情况、速度信息、3 维道路曲面参数和语义信息等。

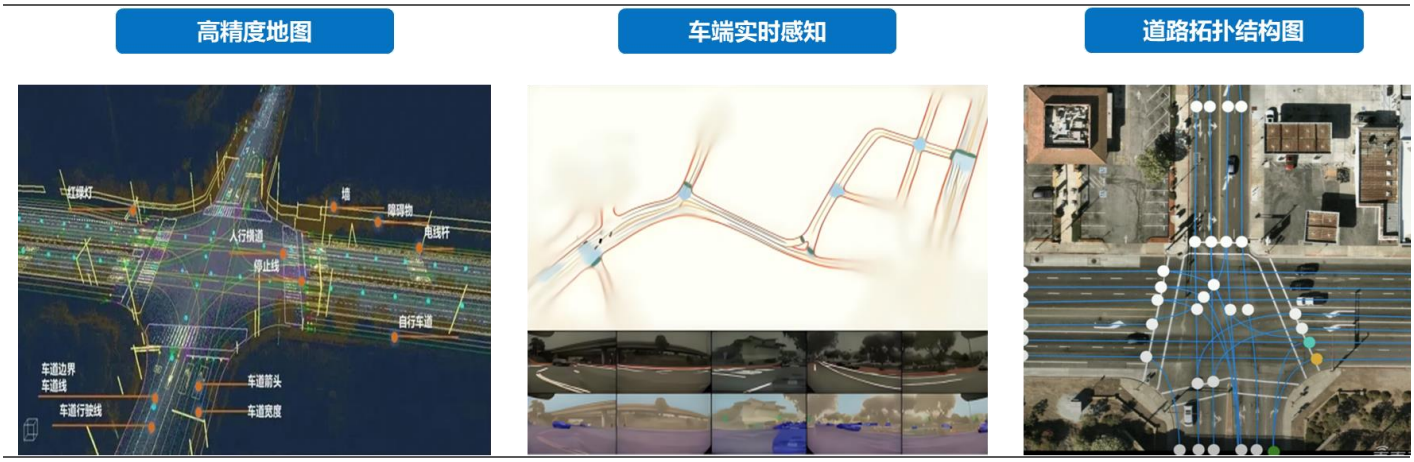
图16. Occupancy 网络架构图



资料来源：特斯拉 AI Day，安信证券研究中心

从 BEV 在线地图升级至矢量地图构建模型 Lanes Network，更有利于下游的规划决策。特斯拉始终坚持无高精度地图的方案，通过车端实时感知+导航地图为下游规划决策提供所需的道路信息，因此**特斯拉在线地图的升级方向就是让其提供的信息密度越来越接近高精度地图。**高精度地图相比于导航地图定位精度明显提升，并且可以提供车道级的信息（车道线的数量、边缘位置），这一点特斯拉在 2021 年通过在 BEV 空间内对车道线进行完整的在线分割和识别已经实现。但除此之外，**高精度地图还可以提供道路拓扑结构，即车道线之间的连接关系，特斯拉将地图模型升级至矢量地图就是为了补足这一信息。**

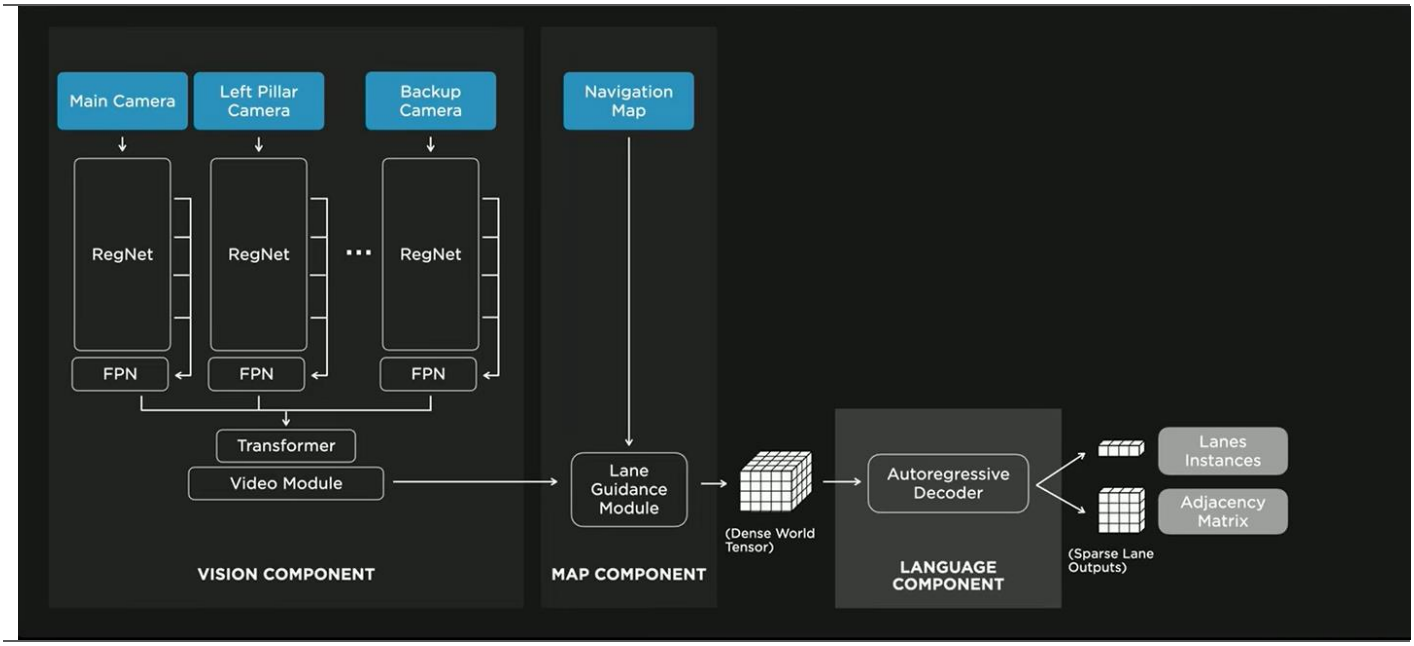
图17. 高精度地图、BEV 地图及矢量地图对比



资料来源：特斯拉 AI Day，安信证券研究中心

特斯拉矢量地图 Lanes Network 包含视觉、地图、语义三个模块，利用 Transformer 生成车道线的关键节点。从网络架构上来说，矢量地图是 Occupancy 感知网络的一个 decoder，将来自感知网络的视觉特征信息、地图的信息整合起来给到语义模块，**这里需要特别注意的是特斯拉所采用的地图是其自己绘制的众包地图，而非高精度地图。**语义模型框架上类似 Transformer 中的 Decoder，将来自视觉模块和地图模块的所有信息进行编码，类似于语言模型中单词 token，再以序列自回归的方式预测节点的位置、属性以及连接关系。

图18. 特斯拉 Lanes Network 模型架构

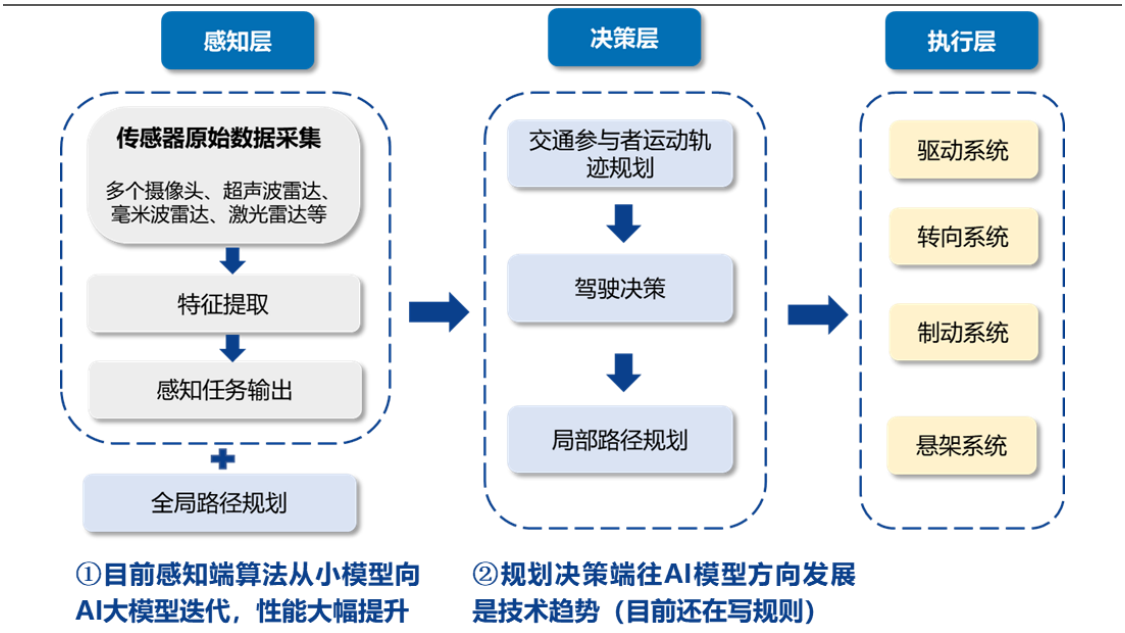


资料来源：特斯拉 AI Day，安信证券研究中心

2.2.4. 2023：规划决策端应用神经网络，实现“端到端”的自动驾驶模型

特斯拉 FSD V12 在规划决策端采用 AI 大模型，更好的处理复杂的交通参与者之间的交互问题，有望实现端到端自动驾驶。在当前自动驾驶模型架构中将驾驶目标划分为感知、决策、执行三个大的模块。目前行业在特斯拉的引领下感知模块均依靠于神经网络实现，但规划决策端依然是基于传统规则，而非神经网络的方式。马斯克在推特上称，特斯拉 FSD V12 将采用“端到端”模型，输入数据是摄像头采集的到的视频流 raw-data，输出数据直接是如方向盘转角多少度的控制决策。可以理解为，除了感知模块，特斯拉在规划决策模块也将采用 AI 大模型、数据驱动的方式来实现。

图19. 自动驾驶系统模块

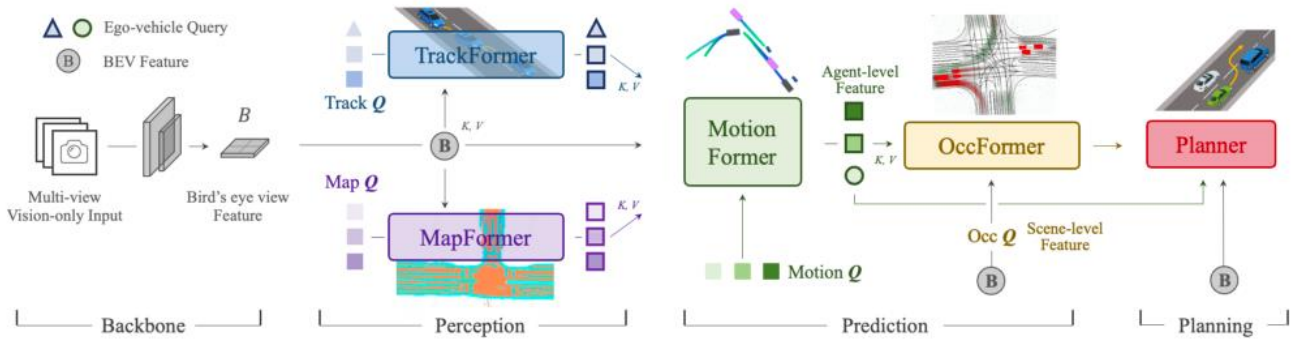


资料来源：安信证券研究中心绘制

规划决策端应用神经网络模型是当前学界、业界共同关注的发展方向。获得 2023 年 CVPR 最佳论文奖的《Planning-oriented Autonomous Driving》提出 UniAD 自动驾驶大模型，UniAD 以“规划”为目标，利用多组 query 实现了全栈 Transformer 的端到端模型。需要

注意的是 UniAD 利用 Transformer 统一了自动驾驶感知、规划决策全栈，但模块之间有明显的区隔，并非完全黑盒，具有一定的可解释性。

图20. UniAD 自动驾驶大模型架构



资料来源：Planning-oriented Autonomous Driving，安信证券研究中心

3. 数据：数据闭环+超算中心造就 FSD 极致迭代速度

如前所述，特斯拉自动驾驶模型从大量的程序员手写规则的 Software1.0 向基于神经网络的 Software2.0 迭代，在 Software2.0 时代下，数据是最为重要的生产资料。我们复盘特斯拉对数据闭环体系的构建可以分为以下两个阶段：1）2016-2019 年：首创影子模式，组建千人标注团队，数据闭环体系初步构建；2）2020 年-至今：逐步发展至 4D 自动标注，数据闭环体系趋于完善，Dojo 超算中心投产进一步提升迭代速度。

3.1. 2016-2019：首创影子模式，数据闭环体系初步构建

特斯拉早在 2016 年首创影子模式，开始在车端大量收集众包数据，2018 年已初步构建了数据闭环体系。一次完整的数据闭环过程分为以下几个步骤：1）从一个初始数据集开始（seed dataset）训练神经网络并部署在车端。2）设计 trigger 触发机制，回传车端收集到的 corner case（如神经网络结果不准确、司机接入接管等）。3）发现这个 corner case 后，写成一个新的 trigger 发送到车端，让车队回传大量的类似数据集。4）对新得到的数据集进行标注，重新训练。**在这一过程中，corner case 的挖掘速度（取决于众包车队的数量以及 Trigger 触发机制的设计）、对类似场景数据的收集速度、数据标注的速度和质量、训练模型的计算资源共同决定了自动驾驶模型的迭代能力。**

图21. 2018 年特斯拉已初步构建数据闭环

资料来源： Tesla Autonomy Day，安信证券研究中心

特斯拉开创影子模式通过大量众包车辆收集 corner case：在有人驾驶状态下，系统包括传感器仍然运行但并不参与车辆控制，只是对决策算法进行验证——系统的算法在“影子模式”下做持续模拟决策，并且把决策与驾驶员的行为进行对比，一旦两者不一致，该场景便被判定为“极端工况”，进而触发数据回传。同时在 2019 年特斯拉已经开始搭建仿真平台，但根据 Tesla Autonomy Day，彼时特斯拉仿真场景存在雨雾等复杂现实环境难以复原等问题，在自动驾驶模型训练中参与度较低。

图22. 特斯拉影子模式

资料来源：Tesla Autonomy Day，安信证券研究中心

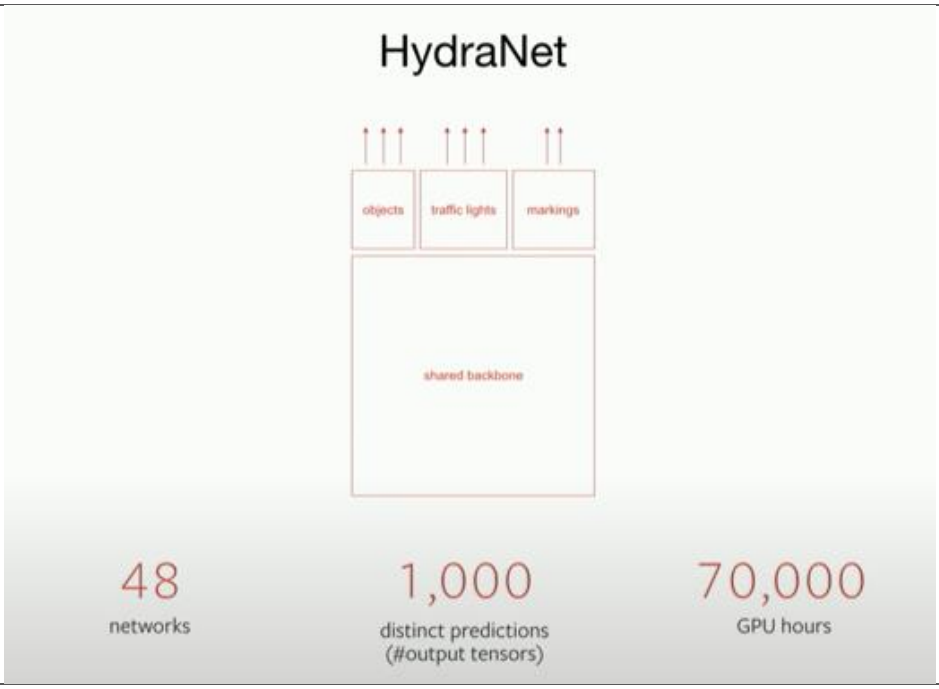
组建超过千人的数据标注团队，保证标注质量。神经网络的训练过程需要给定目标结果（即真值），因此对于收集的数据需要进行标注后才可以用于模型的训练。在 2016-2019 年这个阶段，特斯拉数据标注主要依赖于人工手动标注。特斯拉最早将数据标注外包给第三方团队，但由于外部团队难以及时响应且数据标注质量较低，特斯拉逐步在内部组建了近千人的数据标注团队。同时，在这一阶段训练数据的真值标注是基于 2D 图像的，即在 2D 图像上标注出各种物体（车辆、行人、交通标志等）的位置和类别，形式通常是边界框（Bounding Boxes）。

图23. 基于单个图像的 2D 标注

资料来源：Tesla AI Day，安信证券研究中心

采购英伟达 GPU 自建数据中心，规模尚小，模型单次训练所需时间较长。在特斯拉启动 Dojo 超算中心项目之前，通过采购英伟达 GPU 已初步构建数据中心。根据特斯拉 AI Day 展示资料，2019 年 8 月特斯拉所拥有的 GPU 数量仅约 1500 个，到 2020 年 2 月达到约 1700 个。同时，Andrej 2020 年 2 月在 Scaled ML 会议中的演讲中表示，特斯拉当时自动驾驶模型训练一次需要约 70000 个 GPU 时，因此可以推算在 2019 年末-2020 年初时，特斯拉自动驾驶模型一次训练需要 2 天。

图24. 特斯拉自动驾驶模型单次训练所需约 7 万个 GPU 时（2020 年 2 月）



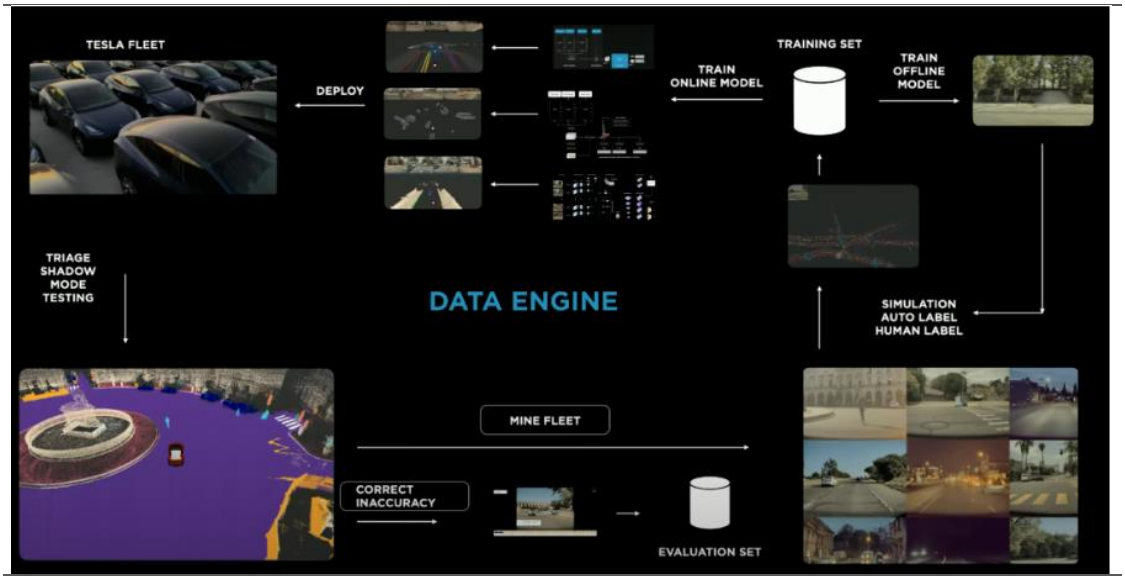
资料来源： Scaled ML，安信证券研究中心

3.2. 2020-2023：逐步升级至 4D 标注，数据闭环体系趋于完善

在初代数据引擎的基础之上，特斯拉升级版数据引擎在标注方案、模拟仿真、云端计算资源三个方面大幅升级，数据闭环系统趋于完善。特斯拉在 2022 年 AI Day 上所展示的数据引擎依然按照模型部署->车端影子模式下发现 corner case 回传至云端->获得大量相似场

景->数据标注后重新训练->再次部署到车端的流程进行，但相比于 2019 年的初代数据引擎版本，主要在标注方案、模拟仿真、云端计算资源三个方面进行了升级。

图25. 特斯拉数据闭环系统（2022）

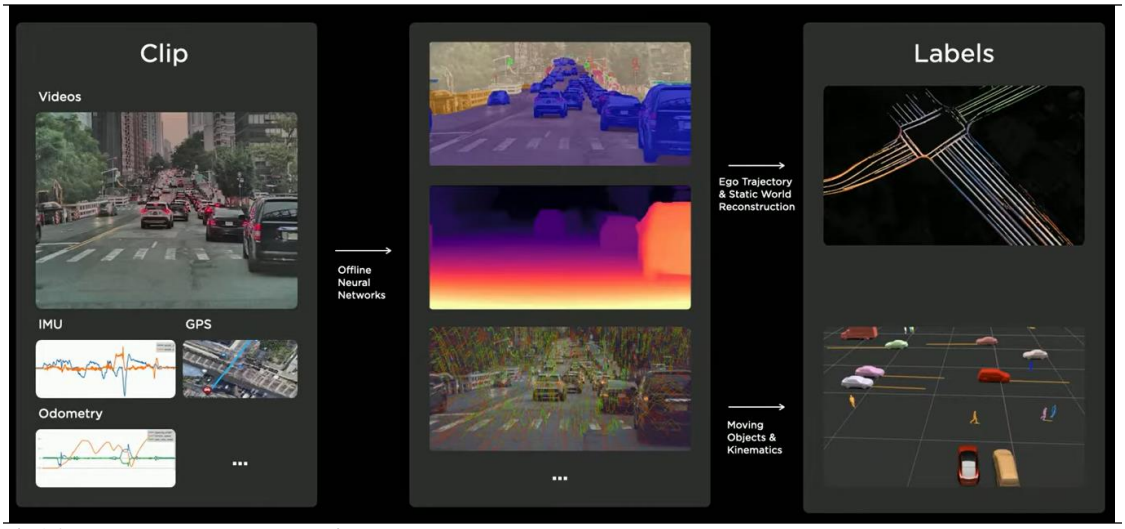


资料来源：Tesla AI Day(2022)，安信证券研究中心

3.2.1. 从 2D 人工标注升级至 4D 自动标注，提升标注效率

从基于图像空间的 2D 标注升级至 BEV 空间下的 4D 标注，大幅提升标注效率。如前所述，对采集的原始数据进行标注来作为神经网络模型的目标结果进行训练。因此训练传统的基于单个摄像头的感知模型，所需要标注的真值仅在 2D 图像空间中完成即可。**而随着感知模型向 BEV 模型迭代，其所需要的真值需要在 BEV 空间内完成标注。特斯拉采用的方法是基于多趟场景重建技术的 4D 自动标注，**具体步骤如下：1）对单个 Clip（Clip 是 Tesla 标注系统的最小标注单位，一个 Clip 通常包含时长为 45 秒到 1min 的路段内所有传感器的数据）使用一个神经网络隐式地对路面建模，得到重建结果；2）将包含相同路段所有的 Clip 进行拼接对齐，完成多趟重建；3）当有新的旅程发生时，就可以进行、几何匹配，得到新旅程车道线的伪真值（pseudolabel）。**特斯拉自动标注系统可以取代 500 万小时的人工作业量，人工仅需要检查补漏。特别需要指出的是，离线自动标注系统同样是大模型，车载感知模型相当于对离线大模型进行蒸馏。**

图26. 特斯拉 4D 自动标注系统



资料来源：Tesla AI Day，安信证券研究中心

图27. 自动标注系统大幅提高标注效率

	image space(2018)	single trip (2019)	top-view(2020)	multi-trip (2021-)
3D label	unknown	manual	aligned	reconstructed
reprojection	<1 pixel	<3 pixel	<7 pixel	<3 pixel
topology	local	up to trajectory	unlimited	up to reconstruction
Labeling / clip	533 hrs	3.5 hrs	<01 hr (avg)	< 0.1 hrs (avg)
compute / clip	not needed	1 hr	2 hrs	0.5 hrs (avg)
scalability	low	medium	high	very high
eng. effort	low	medium	high	very high

资料来源：Tesla AI Day(2022)，安信证券研究中心

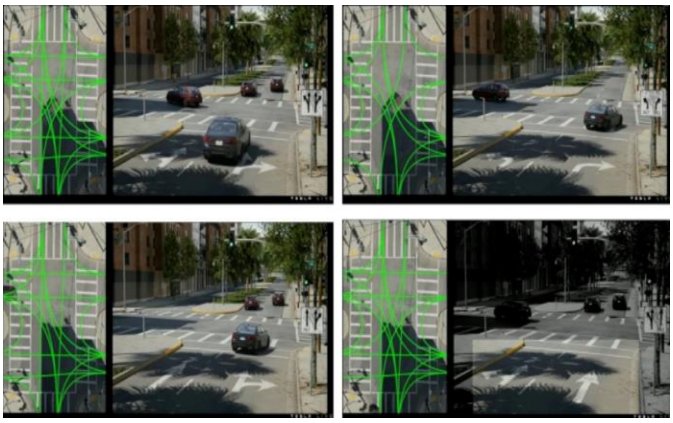
3.2.1. 虚拟仿真技术逐步成熟，赋能模型迭代

加入仿真场景，对所采集的 corner case 进行泛化，提高模型迭代速度。如前所述，在特斯拉初代数据引擎中，在影子模式之下回传 corner case 后，需要再写一个 trigger 发送到车端让众包车队回传类似场景进行训练。但随着模型不断迭代，corner case 出现的概率逐步降低，某些极端场景往往可遇不可求，等待车队回传真实数据耗时较长，在这种情况下，仿真场景是有效的解决方案。特斯拉 Simulation World Creator 具体流程如下：1）由经自动标注的真实场景数据中提取隔离带边界、车道线、道路连接信息等来生成路面网格并进行车道线等渲染；2）植物通过丰富的素材库在路间和路旁随机生成植物房屋等来模拟真实世界中这些物体引起的遮挡效应；3）由导航地图提供信号灯、路牌等其他道路元素；4）加入车辆和行人等动态元素。在这一过程中，通过道路街景随机生成以及车道链接关系的随机生成提高了模型的泛化能力。

图28. 仿真场景中对同一路口生成不同街景进行场景泛化

资料来源：特斯拉 AI Day，安信证券研究中心

图29. 对同一路口生成不同车道关系进行场景泛化



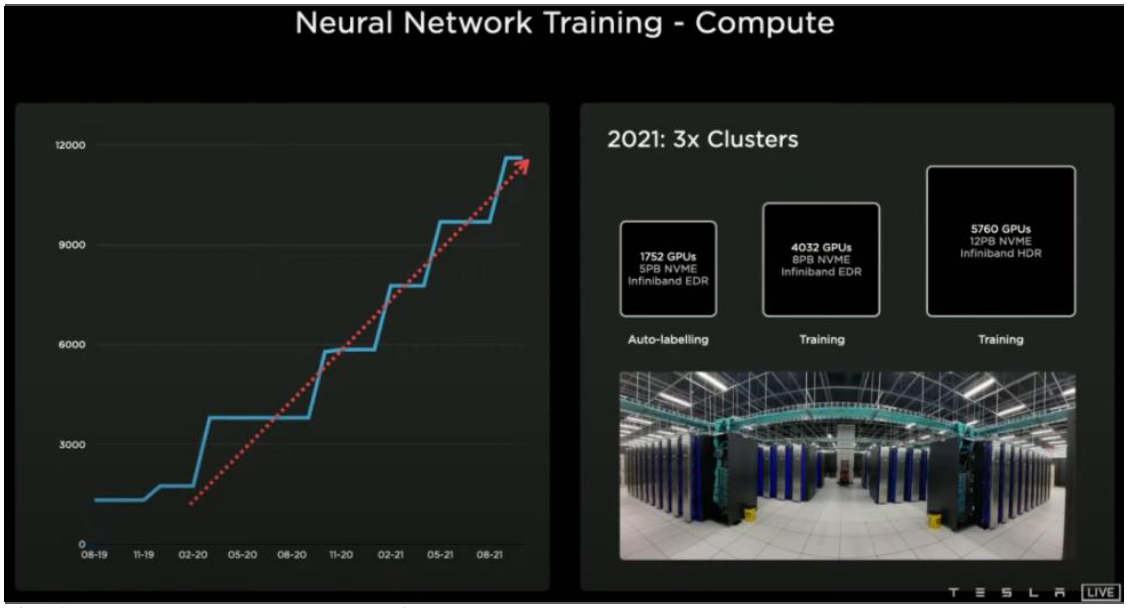
资料来源：特斯拉 AI Day，安信证券研究中心

3.2.1. 云端计算资源不断扩充，Dojo 超算中心正式投产

Dojo 超算中心正式投产，FSD 迭代速度有望进一步大幅提升。根据 Tesla 2021 年 AI Day，自 2019 年以来，特斯拉基于英伟达 GPU 部署的数据中心算力持续提升。2019 年 8 月，特斯拉仅拥有不到 1500 个 GPU，而到了 2021 年 8 月，特斯拉用于云端部署的超级计算机已经拥有 11544 个 GPU。此时，特斯拉具有三个计算集群，其中最大的计算集群具有 5760 个英伟达 A100 GPU（80GB 显存容量），合计 1.8 EFlops 的 AI 算力。而最小的计算集群具有

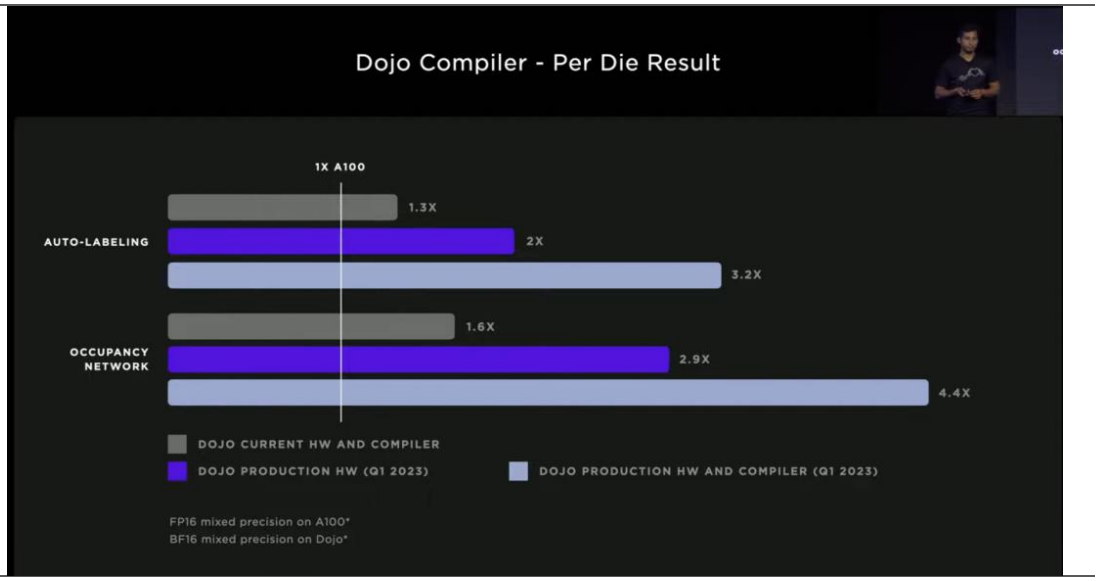
1752 个 GPU 用于自动标注系统。与此同时，特斯拉自 2019 年开始筹备 Dojo 超算中心项目，在 2021 年 AI Day 上正式发布。马斯克表示一方面由于英伟达产能有限，另一方面由于英伟达是通用 GPU，并非针对视频训练的专用芯片，因此特斯拉自研训练芯片可以提高训练效率。根据特斯拉在 AI Day 2022 上公布的数据，与英伟达的 A100 相比，每一颗 D1 芯片（配合特斯拉自研的编译器）在自动标注任务中最高能够实现 3.2 倍的计算性能，而在占用网络任务中最高能够实现 4.4 倍的计算性能。根据 Tesla_AI 的官方推特，Dojo 超算中心已于 2023 年 7 月正式投产，预计 2024 年 2 月达到等效于 10 万个英伟达 A100 的算力，成为全球前五大计算中心。**目前 FSD Beta 的发版速度为平均 20 天一次，我们预计随着 Dojo 超算中心的投产，特斯拉 FSD 的迭代速度会进一步提升。**

图30. 特斯拉基于英伟达 GPU 的超算中心规模持续上升



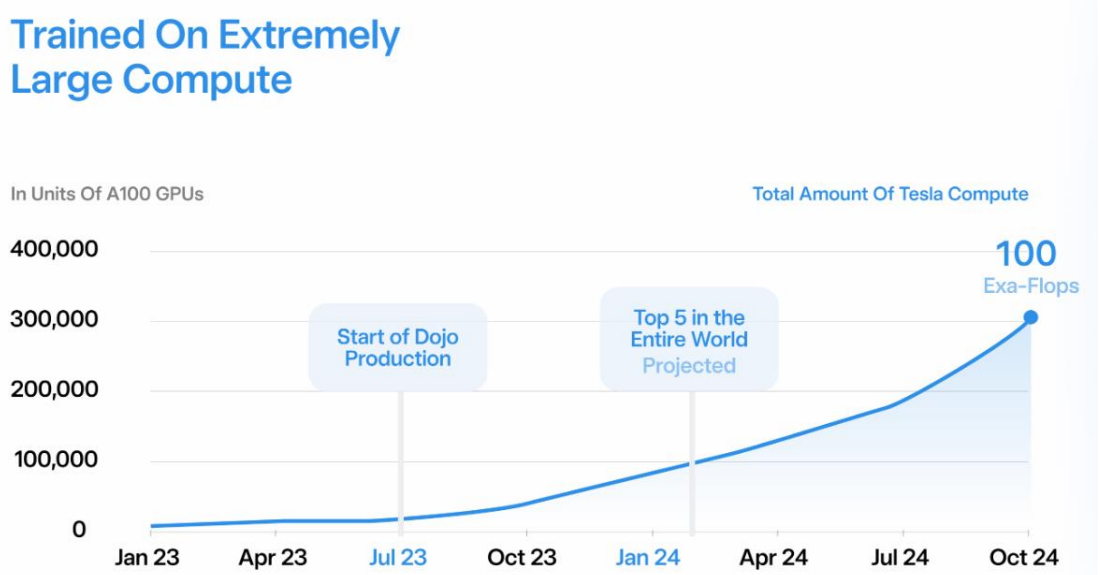
资料来源：Tesla AI Day(2022)，安信证券研究中心

图31. D1 芯片在跑自动标注及占用网络模型的效率高于 A100



资料来源：Tesla AI Day(2022)，安信证券研究中心

图32. 2024 年 2 月 Dojo 超算中心将具备等效于 10 万个英伟达 A100 的算力



资料来源：Tesla_AI 推特，安信证券研究中心

4. 硬件：算力、内存大幅提升赋能自动驾驶算法向大模型迭代

4.1. HW4.0 版本发布，芯片性能、传感器配置全面升级

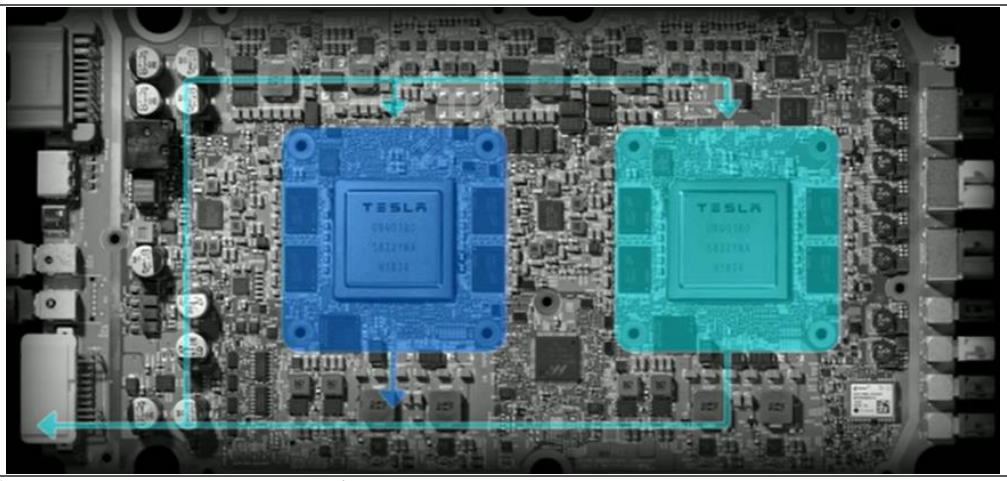
特斯拉自动驾驶硬件自 2014 年逐步从 HW1.0 迭代至 HW4.0，历经核心芯片外采到自研的转变，目前 HW4.0 已开始量产。复盘特斯拉 HW1.0 到 HW4.0 硬件系统配置变化：

Hardware1.0：2014 年 9 月特斯拉推出第一代自动驾驶硬件平台 HW1.0，主芯片采用 Mobileye 的 EyeQ3，同时搭配 Nvidia Tegra3，传感器为 1 颗摄像头+1 颗毫米波雷达+12 颗超声波雷达。特斯拉自始坚持视觉为主的方案，反对使用激光雷达这样的高成本传感器，与同样采用视觉方案的 Mobileye 不谋而合。然而，由于特斯拉和 Mobileye 在数据归属、合作开发模式等方面存在分歧，同时 2016 年 5 月发生的 Autopilot 交通事故成为二者分手的导火索。

Hardware2.0：与 Mobileye 分手后，2016 年 10 月特斯拉基于 Nvidia drive PX2（该平台由 1 颗 Tegra Parker 芯片和 1 颗 Pascal 架构 GPU 芯片构成）推出 HW2.0，算力提升至 12Tops（Mobileye EyeQ3 算力仅 0.256Tops）。传感器方案升级至 8 个摄像头+1 颗前向毫米波雷达+12 颗超声波雷达，这一套传感器配置一直保留至 HW3.0。2017 年 7 月，特斯拉将 HW2.0 升级至 HW2.5，增加了一颗 Tegra Parker 芯片。

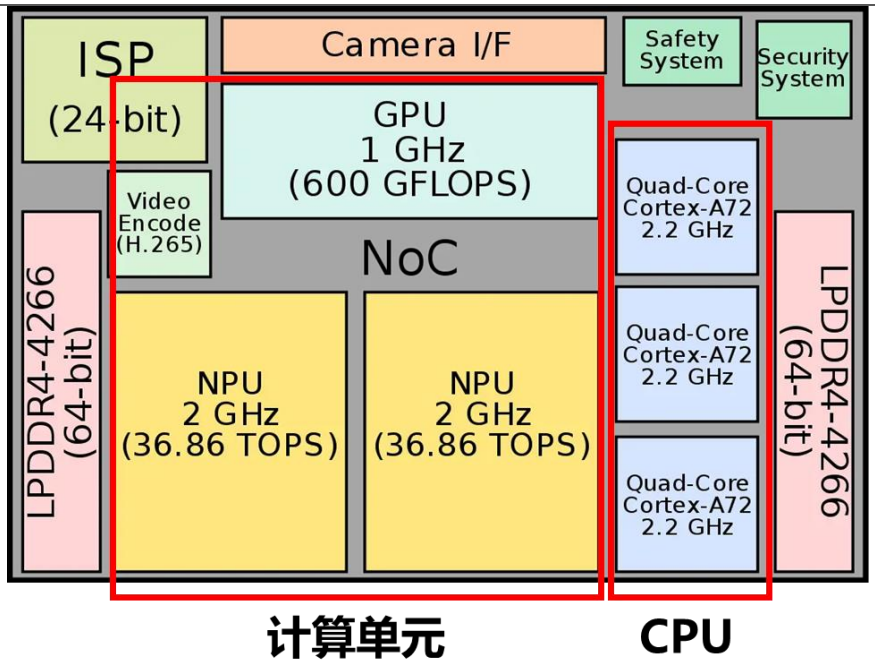
Hardware3.0：特斯拉在与英伟达合作的同时，于 2016 年 2 月开始组建团队自研自动驾驶芯片，历时三年的研发，特斯拉于 2019 年 4 月推出基于 Tesla FSD Computer 的 HW3.0。HW3.0 采用双冗余设计，搭载两块 FSD1.0 芯片，每一块芯片可以独立运算。FSD 芯片采用 CPU+GPU+ASIC 路线：1）CPU：Cortex-A72 架构，共有 12 核，最高运行频率 2.2GHz；2）GPU：最高工作频率为 1 GHz，最高计算能力为 0.6TFLOPS；3）NPU：2 个 Neural Processing Unit（NPU），每个 NPU 可以执行 8 位整数计算，运行频率为 2GHz，单个 NPU 算力 36.86 TOPS，2 个 FSD 芯片的总算力为 144TOPS。

图33. HW3.0 采用双冗余设计



资料来源：Tesla Autonomy Day，安信证券研究中心

图34. 特斯拉 FSD 芯片结构



资料来源：Tesla Autonomy Day，半导体行业观察，安信证券研究中心

Hardware4.0： 2023 年 HW4.0 已搭载于 Model S/X，相较于 HW3.0 在传感器配置、SoC 性能、内存带宽等方面均有大幅提升。**1）传感器配置：**相较于 HW3.0，HW4.0 所搭载的摄像头数量和精度均有所提升。HW4.0 共有 12 个摄像头接口，其中包括 1 个备用、1 个舱内摄像头，实际 10 个摄像头用于自动驾驶感知（其中两个前视），摄像头像素或从 120 万提升至 540 万。此外，HW4.0 预留了 4D 毫米波雷达的接口。**2）SoC 性能提升：**FSD2.0 芯片 CPU 内核由 12 个增加到 20，最大运行频率由 2.2GHz 提高到 2.35GHz。NPU 核从 2 个增加到 3 个（最大运行频率由 2GHz 提高到 2.2GHz），预计域控制器总算力约 400-500Tops.**3）内存方案升级：**从 HW3.0 的 8 颗 LPDDR4 升级至 16 颗 GDDR6，内存容量从 16GB 提升至 32GB，最大内存带宽从 68GB/s 大幅提升至 224GB/s。

表3：特斯拉自动驾驶硬件迭代历程

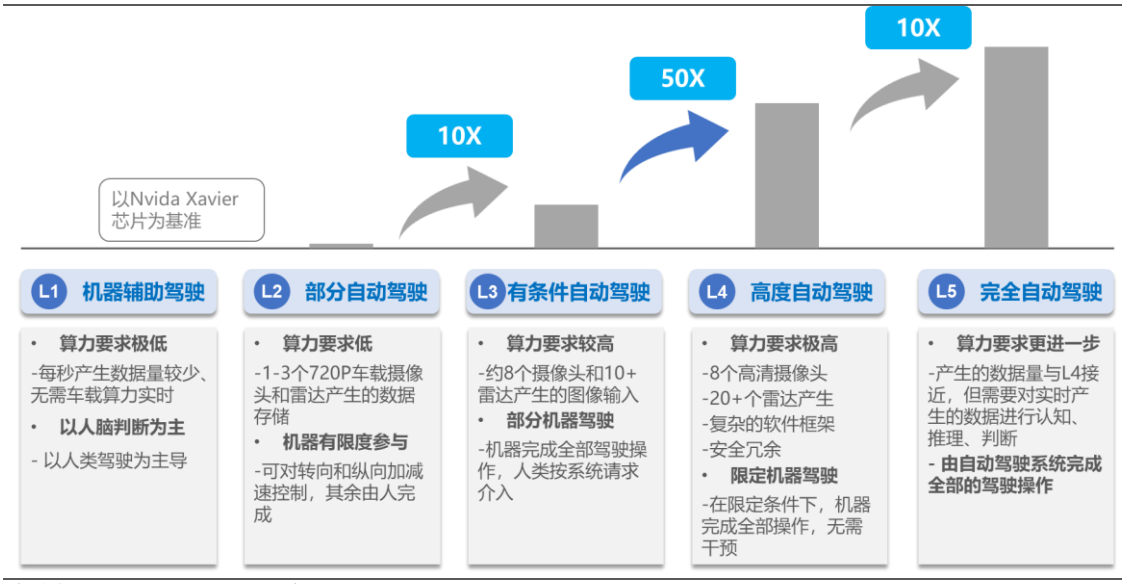
版本			HW1.0	HW2.0	HW2.5	HW3.0	HW4.0
推出时间			2014 年 9 月	2016 年 10 月	2017 年 7 月	2019 年 4 月	2023 年上半年
计算平台			Mobileye EyeQ3	Nvidia drive PX2	Nvidia drive PX2+	Tesla FSD Computer	Tesla FSD Computer
核心处理器			1-Mobileye EyeQ3 1-Nvidia Tegra3	1-Nvidia Parker SoC 1-Nvidia Pascal GPU 1-Infineon TriCore CPU	2-Nvidia Parker SoC 1-Nvidia Pascal GPU 1-Infineon TriCore CPU	2-FSD 1.0（14nm 制程，三星代工，12 个 Cortex-A72 CPU 核+2 个 NPU）	2-FSD2.0 芯片（7nm 制程，三星代工，20 个 CPU 核+3 个 NPU）
FPS 每秒帧数			36	110		2300	-
RAM 内存			256MB	6GB	8GB	8 颗 LPDDR4 总计 16GB（8GB*2）	16 颗 GDDR6，总计 32GB
Flash 存储			-	-	-	4GB*2	三星的 KLUDG8J1ZD，容量 128GB
算力（TOPS）			0.256Tops	12Tops		~144Tops	预估 400-500Tops
硬件配置	摄像头	总数	1 颗 EQ3 系列前视摄像头	8 个摄像头，除了后置摄像头外，其余搭载的摄像头都是使用 Aptina AR0132 摄像头传感器，1/3 英寸、1.2MP CMOS，能够在 60 帧每秒的速度下输出 720p；1280 x 964，每秒 30 帧。			有 12 个摄像头接口，其中一个标注为“备用”，实际摄像头数量很可能是 11 个。预留 1 个+舱内监控 1 个+前视 2 个（提升至 540 万像素）+其他 8 个。
		前置	1	3 个（长焦 250 米、中焦 150 米、广角 60 米），每颗摄像头分辨率为 1280x960 或约 120 万像素。			
		侧置前视	无	2 个（80 米），每颗摄像头的 120 万像素。			
		侧置后视	无	2 个（100 米），每颗摄像头的 120 万像素。			
		后部	倒车用，供应商为 SEMCO（三星电机）	1 个（50 米）。供应商为豪威科技的 OV10635（早期使用的是 OV10630） 720p CMOS 传感器（1160x720，30 帧每秒）。来自日本的 SMK Corporation 是组装该摄像头模块的公司。			
	毫米波雷达		1 个毫米波雷达（160 米）	1 个前置毫米波雷达（160 米）	1 个前置毫米波雷达（170 米）		大概率是 1 个 4D 毫米波雷达
	超声波雷达		12 个中程超声波传感器（5 米）	12 个远程超声波传感器（8 米）			

资料来源：佐思汽研，禹和智能驾驶等，安信证券研究中心

4.2. Transformer 大模型要求自动驾驶芯片具有更强的计算能力

Transformer+BEV 自动驾驶大模型的应用推动车端算力需求提升。车端算力用于量产车上自动驾驶模型推理的过程，可以理解为将训练好的自动驾驶模型部署在车端，自动驾驶汽车实时采集的图像输入到训练好的模型中，依据模型参数算出结果的过程。自动驾驶算法向大模型迭代，参数量大幅提升；同时，随着摄像头精度提升、多传感器融合方案从后融合走向特征级融合，数据量大幅提升，以上因素共同作用使得对车端算力需求提升。根据罗兰贝格的预测，L3 对算力的需求是 L2 的 10 倍。

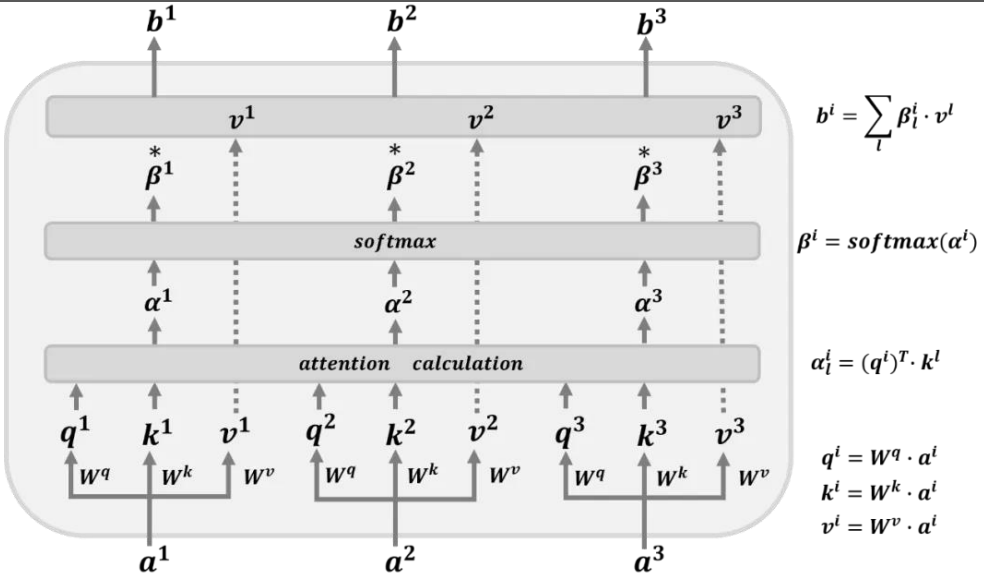
图35. L1 到 L5 级别无人驾驶汽车对算力需求成倍增长



资料来源：罗兰贝格，安信证券研究中心

相比于 CNN，Transformer 模型对芯片浮点计算能力提出更高的要求。传统 AI 芯片主要针对 CNN 模型设计，常使用 INT8 量化操作（将网络中的参数和计算从高精度转换到低精度）以此来减少存储和计算的开销。**CNN 模型中的主要操作是卷积运算和激活函数，对精度的要求较低。**在卷积运算中使用一个小的卷积核在输入图像上滑动并进行元素相乘后相加的运算，如果将输入和卷积核都量化到较低的精度（例如 INT8），在整体的卷积运算中，误差会相互抵消，对最后的结果影响并不大。激活函数通常为分段线性函数，对输入的数值精度同样不敏感。**而 Transformer 模型需要在较高的精度（如 FP16）下进行，要求硬件有高性能的浮点运算能力。**在 Transformer 的核心 Attention 运算中，模型会计算输入序列中每对元素之间的相似度（矩阵乘法），之后通过一个 Softmax 函数转换为权重。这个过程中，点积运算可能会产生非常大或非常小的数值，而 Softmax 函数对这些数值极为敏感，低精度的数值表示可能会导致大量的精度损失。同时，相比于 CNN 只关注局部信息，Transformer 进行全局的信息交换，低精度的数据可能导致误差的累积。除此之外，Transformer 模型的归一化操作中需要计算每个隐藏层的均值和方差，同样需要精确的数值表示。

图36. Attention 工作原理主要为大量矩阵乘法

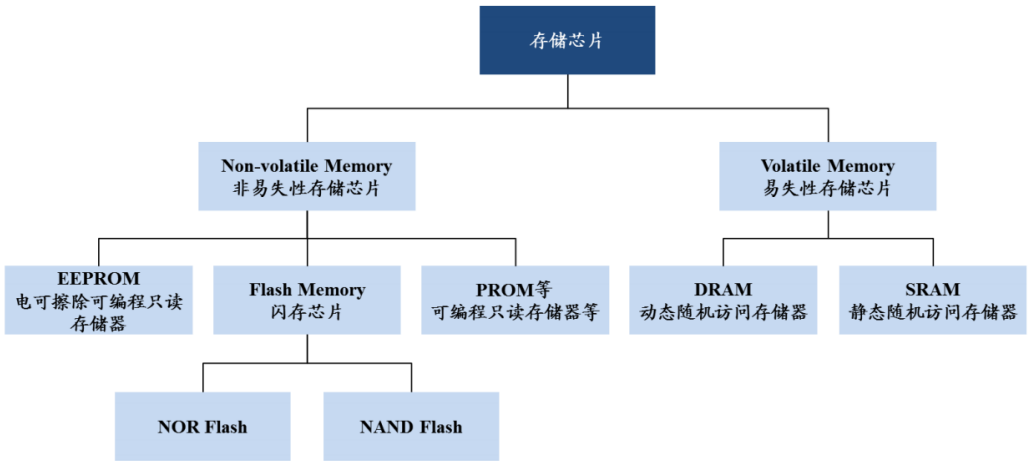


资料来源：知乎@ PaperWeekly，安信证券研究中心

4.3. Transformer 大模型驱动自动驾驶芯片内存方案升级

存储芯片种类较多，主要分为易失性存储器（Volatile Memory）和非易失性存储器（Non-volatile Memory）两大类。根据断电后数据是否丢失，存储芯片可以分为易失型存储和非易失型存储两大类。其中易失性存储器断电后数据丢失，但使用寿命较长且读写速度较快，通常作为 CPU、GPU 等算力芯片的内存，主要包括 SRAM、DRAM 等。虽然 SRAM 带宽较高、存取速度较快，但 SRAM 价格较高，不适合大规模用于车载领域。 DRAM 则分为 DDR、图形 DDR（GDDR）和低功耗移动 DDR（LPDDR）三大类，其中 LPDDR 适合用于对面积和功耗较为敏感的移动和汽车应用。

图37. 存储器的分类



资料来源：聚辰股份招股书，安信证券研究中心

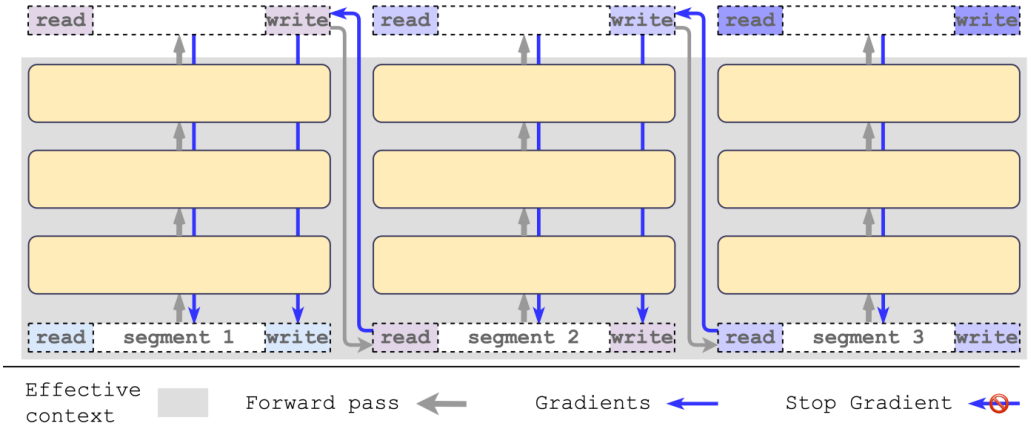
自动驾驶感知模型从 CNN 小模型向 Transformer 大模型迭代过程中，对内存的消耗大幅提升。自动驾驶算法模型对 DRAM 的需求主要来自于三个方面：1）传感器传输的数据，随着摄像头精度的提高，增加内存的需求；2）模型参数（权重矩阵），每一次模型的运算都需要从 DRAM 中加载权重矩阵，模型参数越大，对内存要求越大；3）储存模型计算的中间结果。相比于 CNN，Transformer 模型在以上三个方面均对内存的需求更高。

传感器：单个摄像头的带宽需求=像素数×帧率×颜色深度（每个像素需要多少位来表示），在假设帧率和颜色深度不变的情况下，单个摄像头从 200 万像素升级到 800 万像素，对带宽的要求提升 4 倍。

模型参数：对于深度学习模型来说大部分的空间由参数占据，在车端模型推理过程中的每一次前向传播都需要将模型参数从内存中加载到计算单元中，随着模型参数量的增加，对内存的需求大幅增加。根据佐思汽研数据，传统的目标检测模型尺寸大小通常只有 20MB，而应用于自动驾驶中的 Transformer 模型参数至少在 11 亿以上，即 1.1GB 的权重模型。

储存中间结果：在 Transformer 的自注意力机制中，输入序列的每个元素都需要与其他所有元素进行比较以计算注意力权重。这实际上是在生成一个注意力矩阵，其中第 i 行和第 j 列的元素表示第 i 个元素对第 j 个元素的注意力权重。因此，对于一个含有 n 个元素的输入序列，需要生成一个 n×n 的矩阵来保存这些权重。**这意味着这对针对中间结果的存储空间的需求增长与输入序列的平方成正比。**而对于 CNN 模型，其卷积操作一般只涉及到输入数据的局部区域，所需存储空间相对较小。

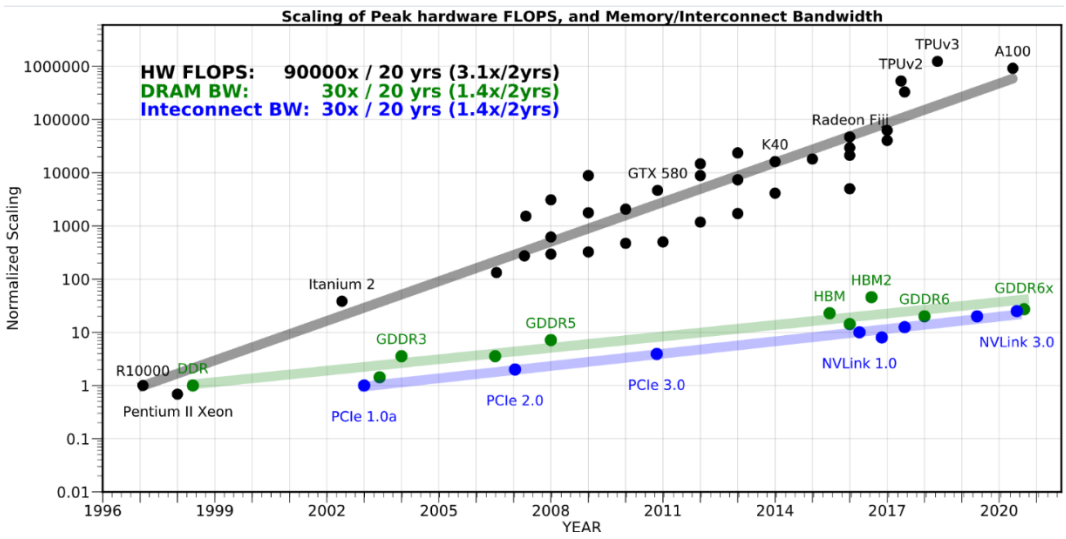
图38. Transformer 每层运算都需要对存储系统进行访问



资料来源：Recurrent Memory Transformer，安信证券研究中心

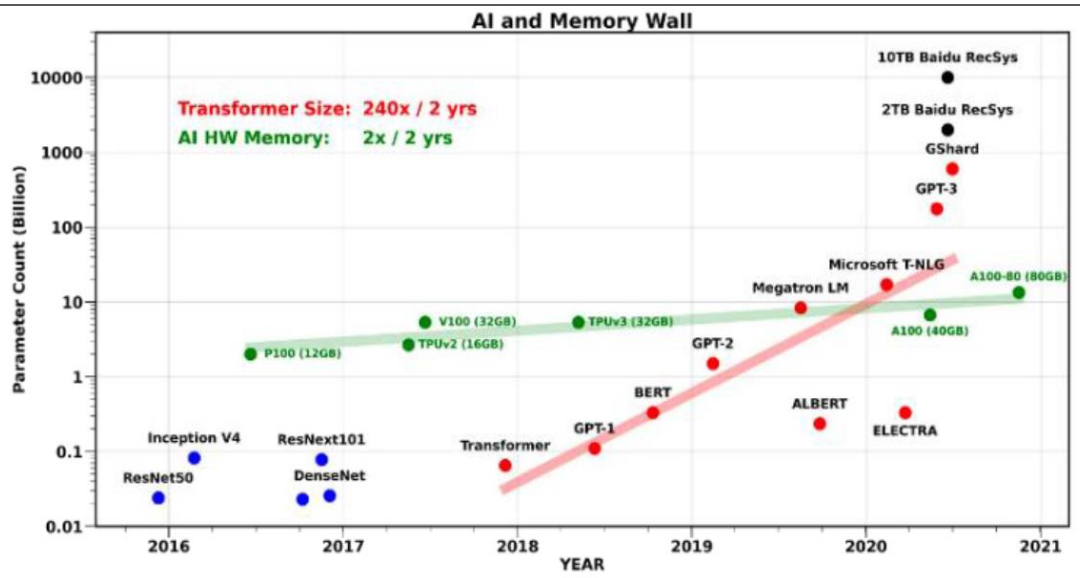
内存的访问速度成为限制芯片有效算力的瓶颈。由于存储器和处理器的工艺不同，二者的性能差距越来越大。存储器受益于制程技术的进步，每 18-24 个月，集成电路上可容纳的晶体管数量就会翻倍；然而，内存的速度提升则主要依赖于电荷存储和访问技术，其进步速度要慢得多。梳理过去 20 年芯片算力及内存参数发现，硬件的峰值计算能力增加了 90,000 倍，但是内存/硬件互连带宽却只是提高了 30 倍。除此之外，根据 UC Berkeley RISELab 数据，大型 Transformer 模型中的参数数量呈指数级增长，每两年增长 240 倍，而单个 GPU 内存仅以每 2 年 2 倍的速度增长。存储器的性能跟不上处理器，导致处理器等待数据的时间远长于运算所消耗的时间，即产生“内存墙”问题。随着自动驾驶感知算法从 CNN 小模型向 Transformer 大模型迭代，参数量大幅提升，内存墙问题愈加明显，成为限制系统性能的瓶颈。

图39. 内存增长速度远低于芯片算力增长速度



资料来源：Github，安信证券研究中心

图40. 内存增长速度远低于模型参数增长速度



资料来源：AI and Memory Wall，安信证券研究中心

为解决内存墙问题，自动驾驶领域所采用的内存类型从 LPDDR4/LPDDR5 向 GDDR6 发展。如前所述，LPDDR 凭借低功耗优势成为自动驾驶领域的主流内存方案。如特斯拉 HW3.0 中搭载 8 颗镁光的 LPDDR4 芯片（单 SoC 配 4 颗），单颗内存容量为 2GB、域控平台总内存容量为 16GB。同时，若按照 LPDDR4 最高频率 4266MHZ 的速率计算，每颗 32 位的位宽，则单 SoC 总传输带宽=4266MHZ(频率)*32(位宽)*4(单 SoC 有四颗 LPDDR4)÷8 = 68.25G/S。随着自动驾驶从 CNN 小模型向 Transformer 大模型迭代，驱动自动驾驶芯片采用更高性能的内存方案，特斯拉 HW4.0 首次将 GDDR6 应用在车载中。GDDR6 是一种用于图形处理器（GPU）和其他高性能计算应用的高带宽内存技术，满足高吞吐量内存的需求。特斯拉 HW4.0 搭载了 16 颗 GDDR6（单 SoC 配备 8 颗），域控平台总内存容量升级至 32GB、单 SoC 对应的理论最大带宽提升至 224GB/s。根据佐思汽车数据，特斯拉 HW4.0 搭载的 16 颗 GDDR6 芯片，总成本约 160 美元；而 HW3.0 搭载 8 颗 LPDDR4 芯片，总成本仅约 28 美元。

表4：HW4.0 相比于 HW3.0 内存方案升级

	HW3.0	HW4.0
内存芯片方案	128-bit LPDDR4	128-bit GDDR6
内存芯片频率	4266MT/s	14000 MT/s
内存芯片数量	8	16
内存容量（域控）	16GB	32GB
内存带宽（单 SoC）	68.3GB/s	224GB/s

资料来源：semianalysis，安信证券研究中心

5. 相关标的

通过复盘了特斯拉自 2014 年至今在自动驾驶领域的核心进展，我们得出以下结论：

1）**算法端：**特斯拉逐步从基于规则的自动驾驶算法向基于神经网络的算法迭代，确立了 BEV+Transformer 的感知范式。通过 AI 大模型赋能，特斯拉在纯视觉的方案下实现了城市 NOA 的大范围落地。特斯拉在感知端通过神经网络大模型解决了城市场景的在线地图问题、一般障碍物识别问题，为行业实现城市 NOA“脱高精度地图”提供了技术上的可行性。**因此，我们认为未来国内主机厂及相关算法供应商均将效仿特斯拉持续加码“重感知、轻地图”的技术路线。2023 年下半年，我们有望初步看到国内主机厂“脱图算法”量产上车，完成脱图之路从 0 到 1 的突破。展望 2025 年，则有望看到城市高阶自动驾驶功能普及至全国多数城市，高阶自动驾驶渗透率开始加速提升。**

表5：2023 年下半年国内脱图城市 NOA 有望大规模落地

主机厂	时间	描述
小鹏	2021 年 1 月 26 日	正式推送高速 NGP
	2022 年 9 月 27 日	广州首发城市 NGP（有高精度地图）
	2023 年 3 月 31 日	G9 和 P7i Max 版车型上城市 NGP 新增开放广州、深圳、上海三地；无图城市能够带来红绿灯识别、启停，以及无车道线的绕行等场景。
	2023 年 6 月 15 日	城市 NGP 在北京开放。
	2023H2	大部分的无图城市都能够有接近城市 NGP 的能力，XNGP 落地 50 城。
	2024	实现全场景（高速+城区+泊车）的领航辅助驾驶，XNGP 落地 200 城。
蔚来	2020 年 10 月	推送高速领航辅助驾驶 NOP
	2023 年 1 月	推送 NOP+Beta，高速领航辅助驾驶体验升级
理想	2021	标配高速 NOA
	2022	AD Max 2.0 视觉融合 Lidar，AD Pro 2.0 纯视觉高速 NOA
	2023Q2	城市 NOA（脱高精度）内测
	2023 年底	城市 NOA（脱高精度）覆盖 100 城
问界 M5 智驾版	2023Q2	华为城区 NCA 落地 5 城（有高精度地图）
	2023Q3	华为城区 NCA 落地 15 城（无高精度地图）
	2023Q4	华为城区 NCA 落地 45 城（无高精度地图）
极狐阿尔法 S	2022 年 9 月	在深圳开通城市 NCA 功能
	2023 年 3 月 21 日	在深圳、上海、广州三城开通城市 NCA 功能
上汽智己	2023 年 4 月	L7 推送 5 个城市高速 NOA，年内推广至全国
	2023 年内	智己城市 NOA 领航辅助以及替代高精地图的数据驱动道路环境感知模型公测。

注：NGP、NOA、NCA、NOP 均指领航辅助驾驶，各主机厂命名有所差异
资料来源：小鹏汽车 2023 技术架构发布会，理想汽车“双能战略”纯电解决方案发布会，2023 华为智能汽车解决方案发布会等，安信证券研究中心

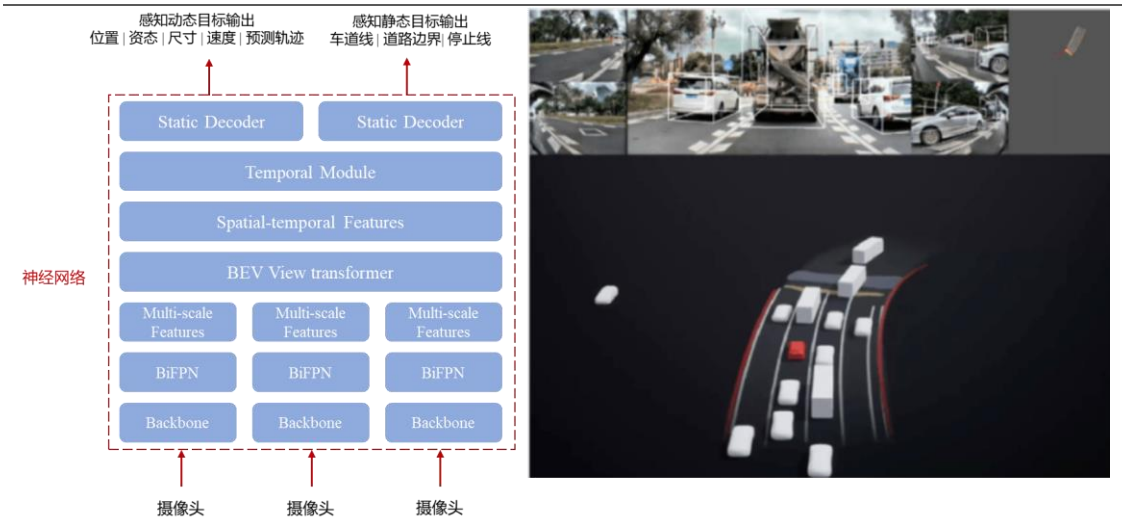
2） **数据：数据闭环+超算中心造就了特斯拉 FSD 极致的迭代速度。**特斯拉早在 2019 年数据引擎已初步构建，2020 年至今在数据标注、仿真引擎、云端计算资源三个方面大幅升级，形成数据飞轮，目前特斯拉 FSD 的发版速度平均为 20 天，随着 Dojo 超算中心的投产，未来 FSD 迭代速度将进一步加快。**因此我们认为，随着自动驾驶进入 AI 大模型时代，数据闭环能力将成为自动驾驶厂商的核心竞争要素。**

3） **硬件端：**特斯拉从 HW3.0 起搭载自研的 FSD 芯片，2023 年 HW4.0 量产，在传感器、芯片性能方面大幅升级。随着自动驾驶算法从小模型向 Transformer 大模型迭代，要求车端算力、内存需求大幅提升。**对于国内自动驾驶厂商而言，通过 Transformer 模型实现城市 NOA 的落地，将更加青睐于大算力 SoC 及高传输带宽存储器。**

5.1. 小鹏汽车

小鹏推出基于 BEV+Transformer 的 XNet 视觉模型架构，自动标注工具+“扶摇”智算中心打造数据闭环。在小鹏 2022 年 1024 科技日中，负责人吴新宙多次强调未来 G9 搭载的 XNGP 自动驾驶系统将无需依赖高精地图，实现城市、高速和地下停车场的全场景应用，其主要的思路是在原有硬件基础上，推出新的视觉感知架构 XNet。其利用多相机多帧和雷达传感器数据的融合算法，直接输出 BEV 视角下交通参与者的静态和动态信息（状态、速度、行为预测等），具备实时生成高精地图的能力。数据闭环方面，小鹏推出的全自动标注系统将效率提升近 45,000 倍，以前 2,000 人年的标注量，现在 16.7 天可以完成。同时，小鹏自动驾驶基础设施建设国内领先，是国内最先布局超算中心的整车厂。2022 年 8 月小鹏汽车成立自动驾驶 AI 智算中心“扶摇”，由小鹏和阿里联合出资打造。据小鹏汽车 CEO 何小鹏介绍，该中心具备 60 亿亿次浮点运算能力（60000TFLOPs），可将自动驾驶算法的模型训练时间提速 170 倍，并且未来还具备 10~100 倍的算力提升空间。目前小鹏基于高精度地图的方案已实现五个城市的城区 NOA 的落地，预计今年年底无图城市 NOA 落地 50 城，明年实现落地 200 城的目标。

图41. XNet 感知架构重感知、轻地图

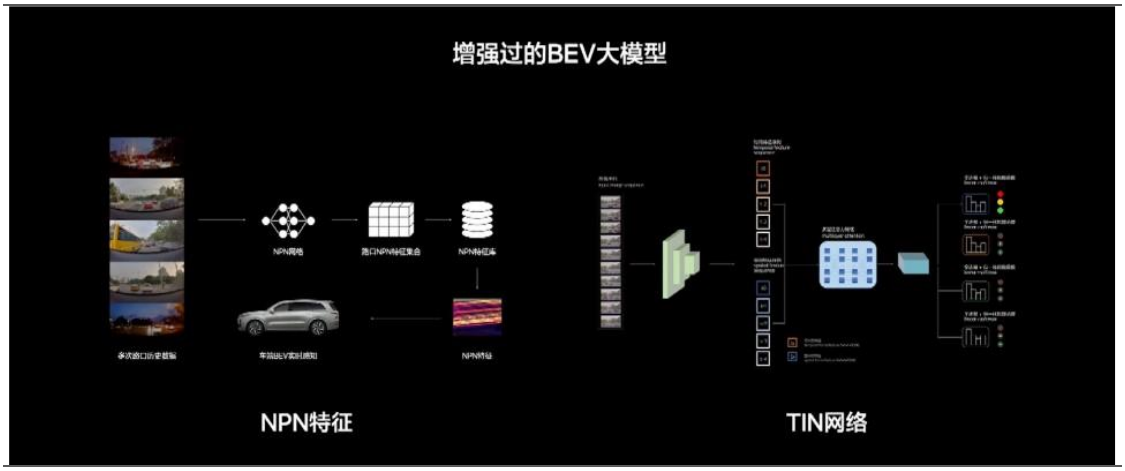


资料来源：小鹏科技日资料，安信证券研究中心

5.2. 理想汽车

理想自研 NPN 特征网络和 TIN 网络，赋能 BEV 大模型实现脱高精度地图。BEV 模型通过将不同视角的摄像头采集到的图片统一转换到上帝视角，相当于车辆实施生成活地图，补足了自动驾驶后续决策所需要的道路拓扑信息，是实现脱图的基础。然而在城市复杂路口场景下，传感器视野易被遮挡、路口跨度大，BEV 模型感知结果易丢失信息，理想采用自研的神经先验网络（NPN 网络）提前提出复杂路口的特征并存储，与 BEV 模型的特征层融合得到更稳定的输出结果。同时，理想通过学习大量人类司机在路口对于信号灯变化的反应训练了端到端的信号灯意图网络（TIN 网络），可以根据输入的视频图像实时给出路口不同同行意图的概率值。

图42. 理想 NPN 先验神经网络和 TIN 信号灯意图网络赋能 BEV 大模型



资料来源：理想家庭科技日，安信证券研究中心

脱图城市 NOA 落地在即，通勤 NOA 解决用户核心痛点。据理想家庭科技日资料，理想城市 NOA 的推送逻辑主要取决于该城市复杂路口 NPN 特征的完成情况，已于 6 月内开启北京和上海的城市 NOA 内测。为解决用户上下班通勤的核心痛点，理想将于 2023 年下半年开放通勤 NOA 功能（即记忆行车）。用户通过设定好通勤路线，通过 1-3 周的训练即可激活，预计可覆盖用户 95%的通勤场景。

5.3. 德赛西威

德赛西威是国内汽车智能化领域的头部 Tier1，深度绑定英伟达这一高算力芯片供应商资源，随着高阶自动驾驶落地对主芯片算力需求的提升将进一步夯实其在自动驾驶域控制器领域的核心地位。公司 2016 年开始前瞻性布局自动驾驶领域业务，2022 年实现 25.7 亿元营收，

同比增长达 83.07%，营收占比 17.22%。目前公司已量产的自动驾驶域控制器包括 IPU01 至 IPU04。其中，IPU01 和 IPU02 主打高性价比方案，主要搭载于中低端车型，实现 L1 和 L2 级别的驾驶辅助功能。根据公司官方公众号，公司基于英伟达 Xavier/Orin 开发的 IPU03 和 IPU04 域控制器主打高算力、高性能，能够满足 L2+的高阶自动驾驶需求，并且已经分别获得小鹏、理想等头部主机厂的量产支持，未来随着 L2+智能车放量自动驾驶域控制器有望为公司持续贡献业绩增量。

6. 风险因素

技术进步不及预期：马斯克称 FSD V12 将采用“端到端”自动驾驶模型，并不再是 Beta 版本。目前行业里在感知端均采用神经网络、数据驱动的方式，但是规划决策端依然以传统算法为主，在规划决策端应用神经网络模型属于学界前沿研究方向，行业里尚无落地先例。特斯拉走在行业前列创造性的采用“端到端”的网络模型面临较大的技术挑战，若技术进步不及预期，将影响 FSD V12 的推送节奏。

非北美地区进展不及预期：目前特斯拉 FSD 功能仅在北美地区推送，在欧洲、亚洲地区 FSD 可能面临禁入风险导致在 FSD 在非北美地区进展不及预期。

商业化进展不及预期：根据 TroyTeslike 统计数据，随着 FSD 价格不断上升，FSD 选装率有所下降，目前 FSD 选装价格已高达 1.5 万美元，若 FSD 价格进一步上涨或 V12 功能体验不及预期，将使得 FSD 商业化进展不及预期。

行业评级体系

收益评级：
领先大市 —— 未来 6 个月的投资收益率领先沪深 300 指数 10%及以上；
同步大市 —— 未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-10%至 10%；
落后大市 —— 未来 6 个月的投资收益率落后沪深 300 指数 10%及以上；
风险评级：
A —— 正常风险，未来 6 个月的投资收益率的波动小于等于沪深 300 指数波动；
B —— 较高风险，未来 6 个月的投资收益率的波动大于沪深 300 指数波动；

分析师声明

本报告署名分析师声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

本公司具备证券投资咨询业务资格的说明

安信证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。

免责声明

本报告仅供安信证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准，如有需要，客户可以向本公司投资顾问进一步咨询。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的估值结果和分析结论是基于所预定的假设，并采用适当的估值方法和模型得出的，由于假设、估值方法和模型均存在一定的局限性，估值结果和分析结论也存在局限性，请谨慎使用。

安信证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

安信证券研究中心

深圳市

地　址：　深圳市福田区福田街道福华一路 19 号安信金融大厦 33 楼

邮　编：　518026

上海市

地　址：　上海市虹口区东大名路 638 号国投大厦 3 层

邮　编：　200080

北京市

地　址：　北京市西城区阜成门北大街 2 号楼国投金融大厦 15 层

邮　编：　100034