AI专题

从特斯拉FSD看人工智能
——端到端模型赋能自动驾驶，机器人引领具身智能

西南证券研究发展中心
海外研究团队 王湘杰
2023年8月

核心观点

- **从特斯拉FSD看AI对自动驾驶的赋能：**
  - **技术端：特斯拉率先提出纯视觉方案，端到端自动驾驶成为新路径。**特斯拉基于对第一性原理的坚持以及对成本的考量，率先实行纯视觉方案，认为自动驾驶可以依靠摄像头实现感知和目标识别，其成本优势也将推动自动驾驶汽车加速实现规模化量产。此外，特斯拉基于Transformer大模型推出端到端自动驾驶方案，构建多任务学习神经网络架构HydraNet，引入特征级融合、占用网络和BEV+Transformer范式。其中，BEV算法有助于将摄像头的2D感知转化为3D视觉，占用网络有助于解决长尾问题，Transformer能够利用注意力机制实现更精准的目标识别，并通过添加时序和空间信息使自动驾驶更接近4D真实世界，推动智驾水平迈上新台阶。目前，以特斯拉FSD为代表的自动驾驶系统表明神经网络算法和AI大模型的赋能已经渗透至智能汽车领域。
  - **商业端：汽车软件化趋势明显，整车价值量有望提升。**随着特斯拉FSD自动驾驶软件的推出，其软件能力已成为差异化卖点，FSD套件的盈利模式采用一次性买断制和按月订阅制，且一次性购买价格经过多轮涨价，目前已提升至15000美元。我们认为，特斯拉在售卖整车的同时还可以售卖自动驾驶服务套件，盈利能力进一步增强。未来，自动驾驶系统在AI技术的赋能下有望持续迭代，单车软件价值逐步增长，推动整车价值量提升，智能汽车软件化趋势明显。
- **从特斯拉FSD看AI对人形机器人上的赋能：**
  - **Optimus沿用FSD底座，有望引领具身智能。**人形机器人与自动驾驶的算法底座本质上均可分为感知层、规划层和控制层，且在硬件设施上有较高的重合度和通用性。特斯拉Optimus同样是基于第一性原理，模拟人体设计，在视觉感知上改进占用网络，在规控上优化运动轨迹，使机器人更好地适应现实世界。我们认为，自动驾驶技术的进步与发展将惠及至人形机器人领域，推动人形机器人迭代提速，引领AI下一代浪潮。
- **投资建议：**建议关注自动驾驶产业链和机器人产业链，其中重点关注具备数据优势、算法优势、且有望在软件端率先进行商业化变现的整车厂商。相关标的：特斯拉(TSLA.O)、小鹏汽车(XPEX.N)等。
- **风险提示：**行业竞争加剧风险，技术发展不及预期风险，商业变现不及预期风险等。

目 录

1.1 行业趋势：以自动驾驶技术为驱动，迈向规模化量产

- **从智能驾驶的发展趋势来看：**行业的上半场以电动化为主，核心驱动力与能源电池紧密相关，技术方向主要为辅助驾驶，市场主要关注技术的验证和特定场景的落地；智能驾驶的**下半场以智能化为主，发力方向主要集中于智能座舱领域和自动驾驶领域**，核心驱动力在于高阶辅助驾驶和自动驾驶技术的创新升级，相关车企逐步聚焦产业化、规模化问题，致力于实现高阶智能汽车的商业化量产。
- **从人工智能带来的变化来看：**我们认为**神经网络算法逐渐对各个产业和领域进行深度赋能。**2022年11月OpenAI推出ChatGPT、2023年3月推出GPT-4，表明大语言模型率先对文本端赋能；当前，特斯拉FSD系统迭代至Beta V11.4版本，在架构上进行重大改进，引入BEV+Transformer范式，推动端到端自动驾驶，表明神经网络的助力已渗透到智能驾驶等领域。
- **随着智驾场景从较为简单的高速场景迈向更加复杂的城市场景，我们认为，在人工智能的赋能下，自动驾驶感知技术的进步将在更多智驾场景下显现优势。**

智能驾驶行业趋势

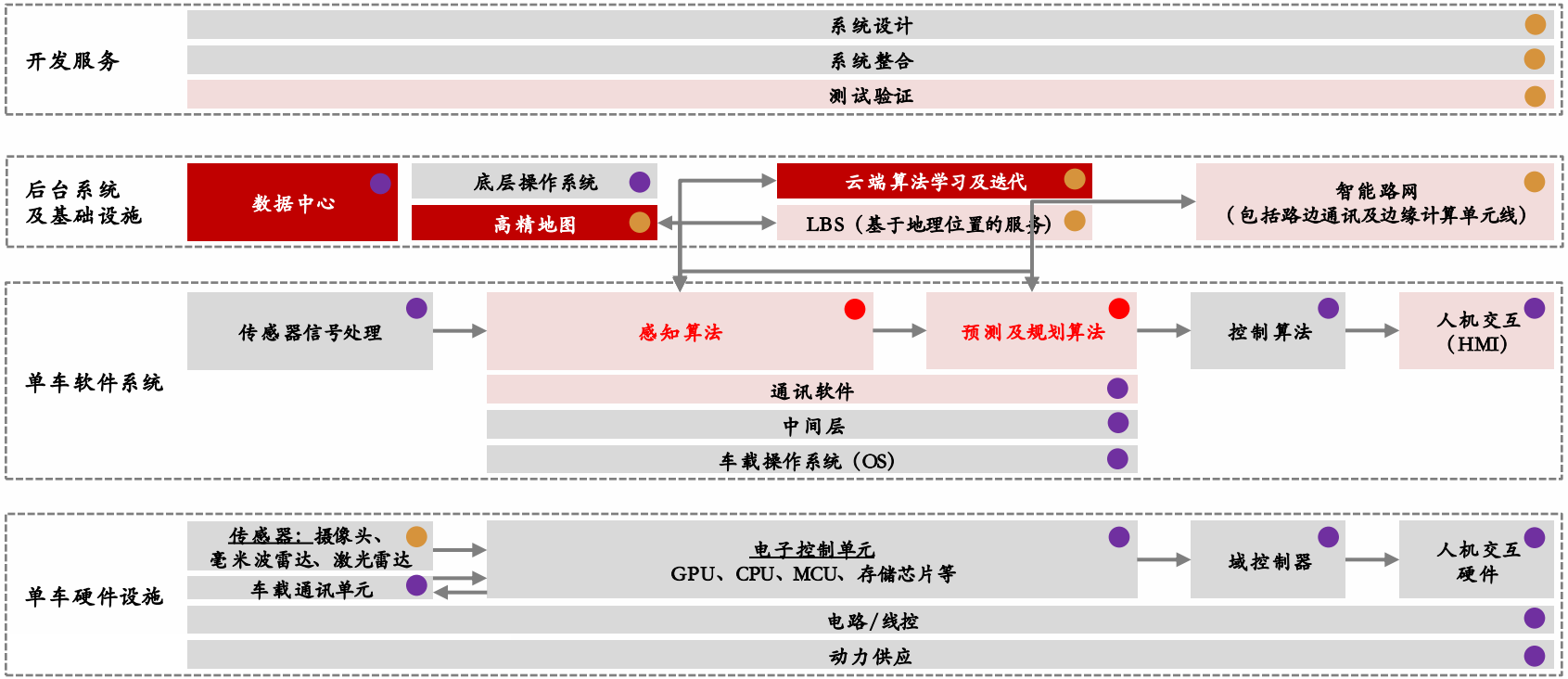
	2016年 电动乘用车渗透率<2%	2022年 电动乘用车渗透率≈20%；智能乘用车渗透率≈5%	2030年 预计智能乘用车渗透率≈50%
行业趋势	**电动化 ——技术验证的上半场**	**智能化 ——商业化量产的下半场**	
行业机会	**动力系统变革：**电动化 **智能驾驶：**L2及以下功能+特定场景L4初试	**电子电气架构变革：**软硬件解耦+中央大算力 **智能驾驶：**多融合方案+L3及以上商业化量产	
竞争要素	技术验证为核心；Demo车场景试点；单一模块的落地能力	商业化量产为核心；满足功能安全的正向开发；规模化/定制化的赋能能力	

资料来源：觉非科技，西南证券整理

1.2 生态圈：算法为核心技术难点，车企与模型厂商探索共建

- **自动驾驶作为行业下半场的重点发力方向之一，其生态圈的构建非常关键。**自动驾驶生态圈可分为四个层级：开发服务、后台系统及基础设施、单车软件系统、单车硬件系统。其中，**软件系统中的感知算法、预测及规划算法是当前的核心技术难点。**我们认为，**自动驾驶解决方案及其生态圈的构建是车企实现产品领先以及差异化体验的核心**，车企可以选择与模型厂商或算法公司合作研发、共同探索，建立基于软件系统和生态圈的核心竞争力。

自动驾驶生态圈及技术发展情况



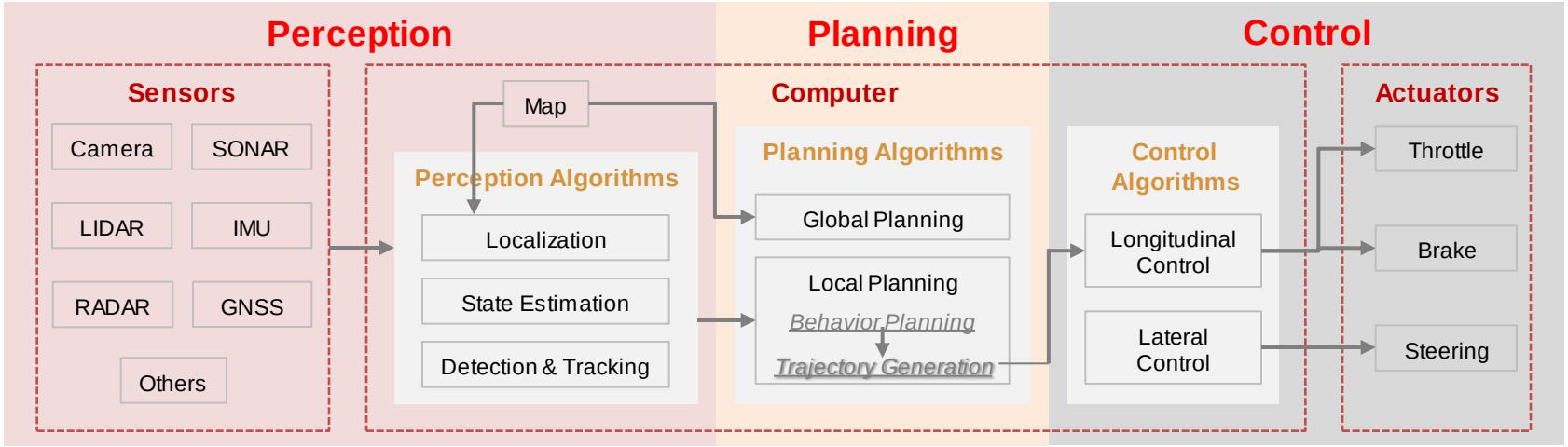
注1：●代表技术完全成熟、●代表技术基本解决、●代表技术核心难点；
注2：粉色方框代表需要中国方案；红色方框代表需要结合客户、车况路况需求深度制定；灰色方框代表基本全球通用。

资料来源：麦肯锡，西南证券整理

1.3 自动驾驶系统：感知是前提，规控决定车辆如何与环境互动

- **自动驾驶系统对应着自动驾驶生态圈中的单车软件系统及部分硬件设施，主要由三个子系统构成：**
  - **感知子系统：感知是规控的前提，由各种传感器和感知算法组成。**车载传感器包括摄像头、激光雷达、雷达、声纳、IMU、GNSS等，用来感知周围环境、监测车辆的定位和状态。感知算法主要包括传感器融合和滤波算法（例如卡尔曼滤波、粒子滤波、贝叶斯滤波），可以帮助减少传感器数据噪声的形成，由此降低测量的不确定性。
  - **规划子系统：利用感知结果，对车辆行为进行最优规划。**车辆采用的最优行为需要通过预测车辆和环境中的其他事物的未来状态来确定，并考虑全局计划、安全性、舒适性及软硬件的约束等。
  - **控制子系统：通过调整车辆的控制元件，准确执行轨迹，实现“实际驾驶”。**控制算法可分为纵向运动控制（例如对车速、与前后车或障碍物距离的控制）和横向运动控制（即垂直于运动方向上的控制，例如转向控制），代表执行器包括油门、刹车和转向等。控制系统决定最终车辆将如何表现并与环境互动。
- **当前，AI对自动驾驶的赋能主要体现在感知环节、以及连接感知和规划的预测环节。**

自动驾驶系统基本架构



资料来源：《CONTROL STRATEGIES FOR AUTONOMOUS VEHICLES》，西南证券整理

1.3.1 感知环节-硬件端：由传感器采集环境信息，主流硬件成本不一

- **感知系统首先需要各种硬件传感器，对周围环境进行感知，并转化为数据和信息。**
- **核心传感器包括：1）视觉摄像头：**价格较低，在百元人民币水平；在目标识别上优势突出；摄像头易受光干扰；对雨、雪、雾等极端天气情况下可能失效。**2）激光雷达：**价格昂贵，当前车载激光雷达单颗售价约数千元甚至上万元人民币；对物体的位置、距离和大小等空间信息的感知更准确；激光雷达基于自发光特性，不受光影响，但易受雨、雪、雾天气的影响。**3）毫米波雷达：**成本较高，在数百人民币至千元区间；不受天气状况限制，环境适应性强；测速，测距能力强。**4）超声波雷达：**价格低，车载超声波雷达单个售价在数十元人民币水平；检测距离短，是泊车功能的重要传感器。

四类自动驾驶传感器性能对比

指标	激光雷达	毫米波雷达	超声波雷达	摄像头
探测距离	＜150m	＞150m	＜10m	＜50m
分辨率	>1mm	10mm	差	差
方向性	能达到1度	最小2度	90度	由镜头决定
响应时间	快（10ms）	快（1ms）	慢（1s左右）	一般（100ms）
精度	极高	较高	高	一般
温度稳定性	好	好	一般	一般
湿度稳定性	差	好	差	差
恶劣天气适应性	差	强	差	差
穿透力	强	强	强	差
成本	高	较高	低	一般
功能	实时建立周边环境的三维模型	自适应巡航、自动紧急制动	倒车提醒、自动泊车	车道偏离预警、前向碰撞预警、交通标志识别、全景泊车、驾驶员注意力监测
优点	精度极高、扫描周边环境实时建立三维模型的功能强大	不受天气影响、探测距离远、精度高	成本低、近距离测量精度高	成本低、可识别行人和交通标志
缺点	成本高、精度会受恶劣天气影响	成本高、难以识别行人	只可探测近距	依赖光线、极端天气可能失效、难以精确测距

资料来源：曼孚科技，西南证券整理

1.3.1 感知环节-硬件端：车企配置各异，4D毫米波雷或成为新标配

- **自动驾驶传感器配置可分为三大派别：1）摄像头派：**以特斯拉为代表，主要通过视觉摄像头模拟人眼视力；**2）激光雷达派：**增加多颗激光雷达，价格较为昂贵，从目前多款车型的传感器配置来看，激光雷达已具备上车能力；**3）毫米波雷达派：**以蔚来、小鹏为代表，采用“多颗毫米波雷达+摄像头”的硬件配置，截至2023H1，部分新势力车企和传统车企已将4D毫米波雷达应用至旗下新晋品牌的主推车型上。

特斯拉及新势力厂商硬件配置情况

汽车厂商	车型	硬件配置				
		摄像头	激光雷达	毫米波雷达	超声波雷达	芯片
特斯拉	model 3/s/x/y	8	0	1	12	特斯拉FSD
理想汽车	L7/L8 Air/Pro	10	0	1	12	1*地平线征程5
	L7/L8/L9 Max	11	1	1	12	2*英伟达Orin-X
蔚来	ET/ES/EC	12	1	5	12	4*英伟达Orin-X
小鹏汽车	G9 570Plus/570Pro/702Pro/650Pro	11	0	5	12	1*英伟达Orin-X
	G9 570Max/702Max/650Max	11	2	5	12	2*英伟达Orin-X

资料来源：特斯拉官网，理想汽车官网，蔚来官网，小鹏汽车官网，西南证券整理

2023H1部分主机厂激光雷达和4D毫米波雷达搭载情况

主机厂	蔚来	小鹏	理想	上汽	路特斯	长安	北汽极狐	高合	广汽埃安	合创	一汽红旗
车型	ES8/ES7/ET7/ET8等	G6/G9等	L7/L8/L9	智己L7/飞凡F7/飞凡	Eletre	深蓝SL03	极狐阿尔法	HI-Phi Y/Hi-Phi Z	LX Plus/Hyper GT	V09	E001
激光雷达	√	√	√	√	√		√	√	√	√	√
4D毫米波雷达				√	√	√					

资料来源：亿欧智库，西南证券整理

1.3.1 感知环节-硬件端：车企配置各异，4D毫米波雷或成为新标配

- **4D毫米波雷达优势凸显，有望成为感知硬件配置新方案。**毫米波雷达是指工作波长介于1-10mm的电磁波雷达，通过向障碍物发射毫米电磁波并接收回波来精确探测物体的**距离、速度、方位**，而**4D毫米波雷达**除探测“距离、速度、方位”外，还可以用于测量**高度**，从而实现四个维度的感知，并具有广视角、高精度、高分辨率等优势，有助于进一步拓展自动驾驶的感知能力。对比其他传感器，毫米波雷达是基于电磁场原理，而激光雷达和摄像头本质上均基于光波原理，均不能在雨雪雾霾等恶劣天气情况下正常工作，而**毫米波雷达可以全天候不受光线和气候的影响，可为其他传感器提供更多冗余**；此外，激光雷达成本高，在一定程度上阻碍了其作为感知装置的硬需求，但得益于其分辨率较高，因此能为车企在开发样车阶段能够提供很好的起点，**若4D毫米波雷达同样具备较高的分辨率，将成为车企更经济的配置选择。**

激光雷达和4D毫米波雷达的性能及应用对比

	相同点	不同点
性能	**①两者都能提供目标物的距离、高度等信息：**4D毫米波在传统毫米波之上添加高度维度，所以其能够探测目标物的高度信息，同时延续传统毫米波的探距能力和测速能力	**①点云稠密度不同：**4D毫米波雷达的点云可能只达到10万点云左右，仅能达到64线以下激光雷达点云数量，而无法达到百线级别的激光雷达点云数量，如128线的产品可能达到140万左右点云
	②两者都能输出三维图像信息：4D毫米波的点云与激光雷达点云类似，具备三维建模能力，能够识别出目标物的三维信息，可区分行人和物体	**②角分辨率水平不同：**4D毫米波雷达的水平和垂直角分辨率可达到1°~2°，而激光雷达的水平和垂直角分辨率可达到0.1°~0.2°
应用	**①算法应用不成熟：**4D毫米波雷达和激光雷达的算法应用都尚未达到到成熟应用阶段，甚至4D毫米波雷达成熟度比激光雷达更低	**①环境适应能力不同：**4D毫米波雷达是全天候传感器，不受雨雪雾尘影响，但激光雷达易受到雨雪雾尘的影响
		②目标物穿透能力不同：4D毫米波雷达可以穿透物体，检测到前车前方的目标物，但激光雷达的点云无法穿透前方目标物

资料来源：亿欧智库，西南证券整理

1.3.1 感知环节-硬件端：传感器技术迭代，有望覆盖更多长尾场景

- **随着各类传感器的迭代升级，更多的长尾场景有望被覆盖，自动驾驶系统的鲁棒性及行车安全将进一步提升。**
  - **激光雷达：具备较好的路面小物体识别能力和鬼探头的识别能力。**
  - **4D毫米波雷达：**在传统毫米波雷达的基础上进行技术升级，**能够识别路面的小物体，也能够识别前车前方的车辆刹车，且识别精度高**，可提前让自动驾驶系统采取制动措施，避免追尾事故。

各类传感器的长尾场景覆盖情况

一级分类	二级分类	长尾场景	激光雷达	4D毫米波雷达	毫米波雷达	摄像头
传感器层	硬件级	传感器表面是否易污渍	表面易积污渍，容易产生噪点	不受影响	不受影响	表面易污渍
		耐高低温环境能力	-40°~85°	-40°~85°	-40°~85°	-40°~80°
	物理级	受目标物表面材质和颜色的影响	能量易被黑色表面吸收	对金属表面目标物易敏感	对金属表面目标物易敏感	不易识别白色物体
内容层	域级	受极端天气的影响	受影响	不受影响	不受影响	影响较小
		进出隧道光线明暗突变	不受影响	不受影响	不受影响	受影响
	目标级	小物体识别能力	可识别	可识别	不易识别	不易识别
		动静物体识别能力	可识别	可识别	不易识别	可识别
	场景级	路边行人和车辆的区分	可区分	可区分	无法区分太近的物体	有条件区分
时域层	超预期级	鬼探头识别能力	可识别	不易识别	不易识别	不易识别
		前车前方车辆的刹车识别能力	不易识别	可识别且精度高	可识别但置信度低，结果易漏检	不易识别

资料来源：亿欧智库，西南证券整理

1.3.2 感知环节-融合方案：纯视觉VS多传感器融合→降本VS安全

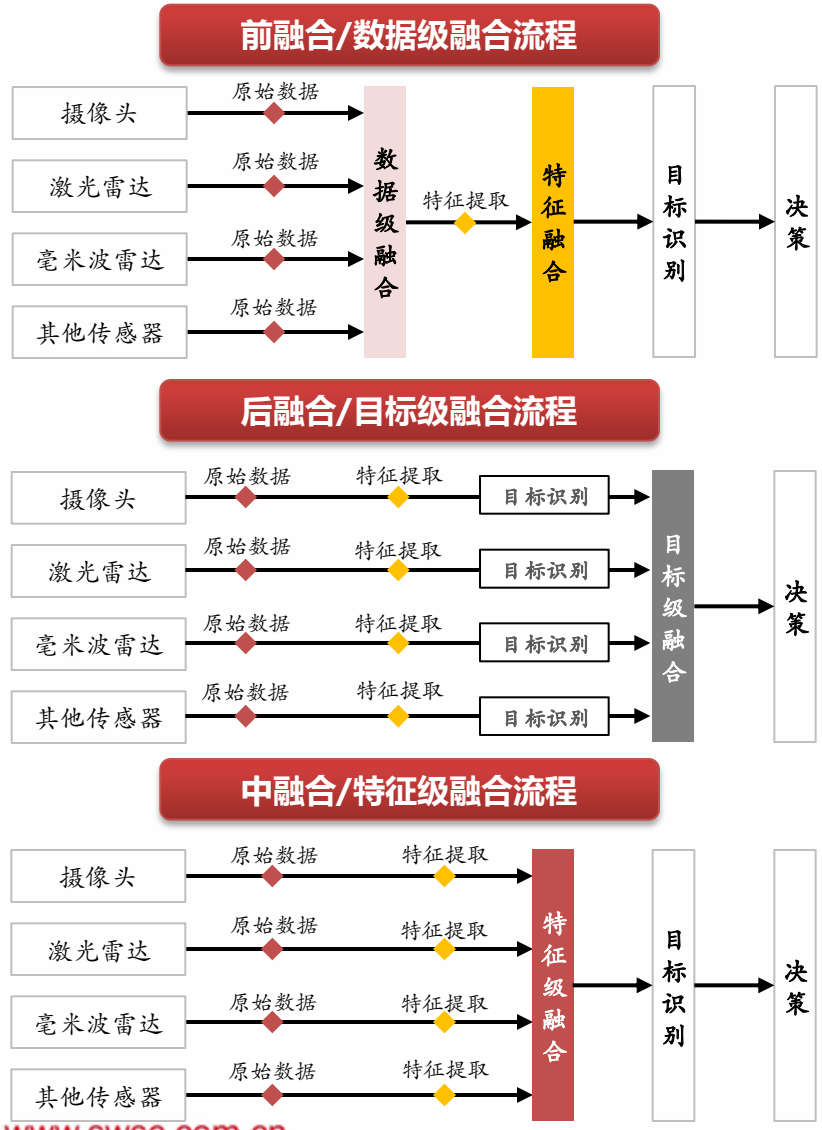
- **从传感器的硬件配置来看，硬件的不同对应着不同的感知路径。**1）纯视觉路线：从第一性原理出发，人类驾驶通过视觉感官识别周围环境，依靠的是一种近乎无意识的感知，因此自动驾驶同样可以通过类似于人眼的摄像头实现识别功能。2）多传感器融合路线：多传感器融合方案除了采用摄像头以外，还采用激光雷达等其他传感器收集车辆周边信息，系统将来自多个传感器的信息和数据在一定的准则下加以分析，为规划和决策提供依据。
- **厂商基于安全考虑采用多传感器融合方案。**在行业发展前期，车企为尽快完成自动驾驶布局，多采用硬件堆料以实现更多功能，尽管目前自动驾驶技术已取得明显进步，但在不能完全保证安全的情况下，多数汽车厂商依然保持传感器的冗余策略。多传感器融合方案可以提升对有些场景的感知精度，同时在某一传感器失效时，其他传感器的数据可以相应补充，但其“融合”技术是该方案当前面临的主要挑战。
- **降本成为车企转向纯视觉方案的现实因素。**事实上，堆料并不意味着性能的绝对提升。此外，在价格战的压力下，车企为节约成本，希望通过精简相关零部件以降低硬件开支，因此，部分厂商逐步去掉价格昂贵的激光雷达，转变为纯视觉方案，尽管该方案成本更低，但其对数据和算法的要求更高，汽车厂商需加大在软件端的投入和布局。

纯视觉感知与多传感器融合感知对比

感知策略	纯视觉方案	多传感器融合方案
信息丰富度	仅有图像语义	包含图像语义+三维点云
三维深度	无	有稀疏的点云深度信息
测距精度	低	高
相对场景可靠性	低可视场景下性能下降	抗干扰能力强
相对成本	低	高
多模态数据融合	无需	数据融合复杂，可能存在冲突
通用模型开发	视觉通用大模型	不同模态数据的模型，算法尚未统一
技术难点	从2D图像重建3D场景	多模态数据融合和对齐

资料来源：CICV，西南证券整理

1.3.2 感知环节-融合方案：后融合为当前主流，中融合为发展趋势



- **从传感器的融合流程上看**：主要前融合（数据级融合）、中融合（特征级融合）、后融合（目标级融合），其中，前融合业内采用较少，**后融合为当前主流，特征级融合有望成为未来发展趋势。**
- **前融合：1)原理：**采集各传感器数据，经过数据同步后，对原始数据进行融合。**2)优点：**保留数据关联性，数据损失少。**3）缺点：**①异构传感器坐标系不一致会导致融合效果不理想，对融合策略要求高。②前融合需要处理大量数据，对算力要求较高。
- **后融合：1）原理：**各传感器针对目标物体单独进行深度学习模型推理，从而各自输出带有传感器自身属性的结果，并在目标层进行融合。**2）优点：**不同传感器独立进行目标识别，解耦性好，各传感器可以互为冗余备份。**3）缺点：**①各传感器经过目标识别再进行融合时，中间损失很多有效信息，影响感知精度；②最终的融合算法是一种基于规则的方法，需根据先验知识来设定传感器的置信度。
- **中融合/特征级融合**从原理上看，该融合方案先将各个传感器通过神经网络提取中间层特征（有效特征），再对多种传感器的有效特征进行融合，从而**更接近最佳推理。**此外，特征级融合相对后融合**数据损失更少**、相对前融合**算力消耗更少，因此，自动驾驶感知融合方案逐步朝特征级融合发展。**

资料来源：九章智驾，西南证券整理

1.3.3 感知环节-视觉表达：BEV实现动态还原，占用网络展现4D泛化世界

- **AI算法在视觉呈现上赋予感知系统“脑补”能力，感知系统逐步具备实时性、更稳定、更精准。**自动驾驶感知系统形成的视觉表达从透视图逐渐发展到鸟瞰图和占用网络，道路还原从2D空间扩展为3D、4D空间，使车辆在动态运动的过程中能够实时构建现实地图，在多颗摄像头的感知下迅速追踪物体的距离和速度、发现被遮挡的物体，并增强现实世界的还原细节和精度，让系统的感知呈现更加符合人类驾驶的需求。

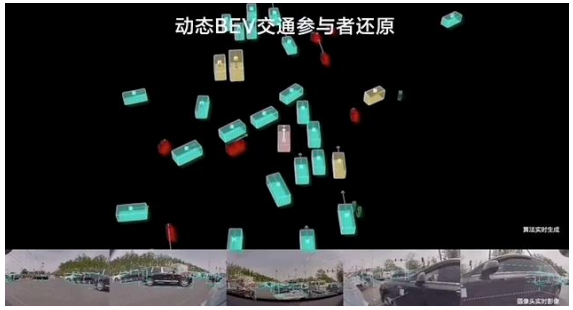
- **透视图（Perspective View）：**即人眼通常看到的2D视图。在人类视觉中，难以看到被遮挡的物体，但在实际驾驶过程中，人类驾驶员可以凭借经验和记忆对可能存在遮挡情况的风险进行规避，但自动驾驶系统如果是基于透视图的视觉进行感知和预测，车辆则很难做到提前预警和规避。

- **鸟瞰图（Birds View）：**即自上而下的视图，具备上帝视角。鸟瞰图感知方案可以在3D空间上分离所有对象，解决透视图中视野被遮挡的问题，减少对自动驾驶对高精地图的依赖，但在高度检测上效果不够理想。

- **占用网络（Occupancy Network）：**占用网络通过算法对物理世界进行**数据化和泛化建模**，在3D空间上测出不同物体的**高度**，呈现**4D视觉**。例如，识别道路上的垃圾桶、临时施工牌等障碍物。

鸟瞰图和占用网络的视觉呈现对比

- 静态BEV网络通过感知还原道路结构，减少对高精地图的依赖。



- 可以解决视野被遮挡的问题，并实时动态还原现实道路的情况

- 可测量出障碍物的高度，识别细节物体

资料来源：2023上海车展·理想汽车，西南证券整理

1.4.1 技术路径：大模型成就端到端自动驾驶，推动感知决策一体化

- **目前，自动驾驶系统的设计主要分为两大技术路径：模块化方案和端到端方案。**两大路径可优劣互补，以上路径当前均在积极探索、相互结合。
- **模块化路径：**涉及众多模块，每个独立的模块负责单独的子任务，例如自动驾驶系统的一级模块可分为感知、规划和控制，每个一级模块下又分为众多子模块，每个模块可基于不同的规则或算法。由于每个独立模块负责单独的子任务，因此出现问题时可及时回溯，并易于调试，具有较强的解释性。
- **端到端路径：**端到端（End-to-End）概念来源于深度学习，端到端路线是指AI模型只要输入原始数据就可以输出最终结果。在自动驾驶的应用中，端到端模型可以将感知、规划和控制环节一体化，通过将车载传感器采集到的信息直接输入**神经网络**，经过处理后直接输出自动驾驶的驾驶命令，潜在性能更佳、优化效率更高。

模块化自动驾驶 VS 端到端自动驾驶

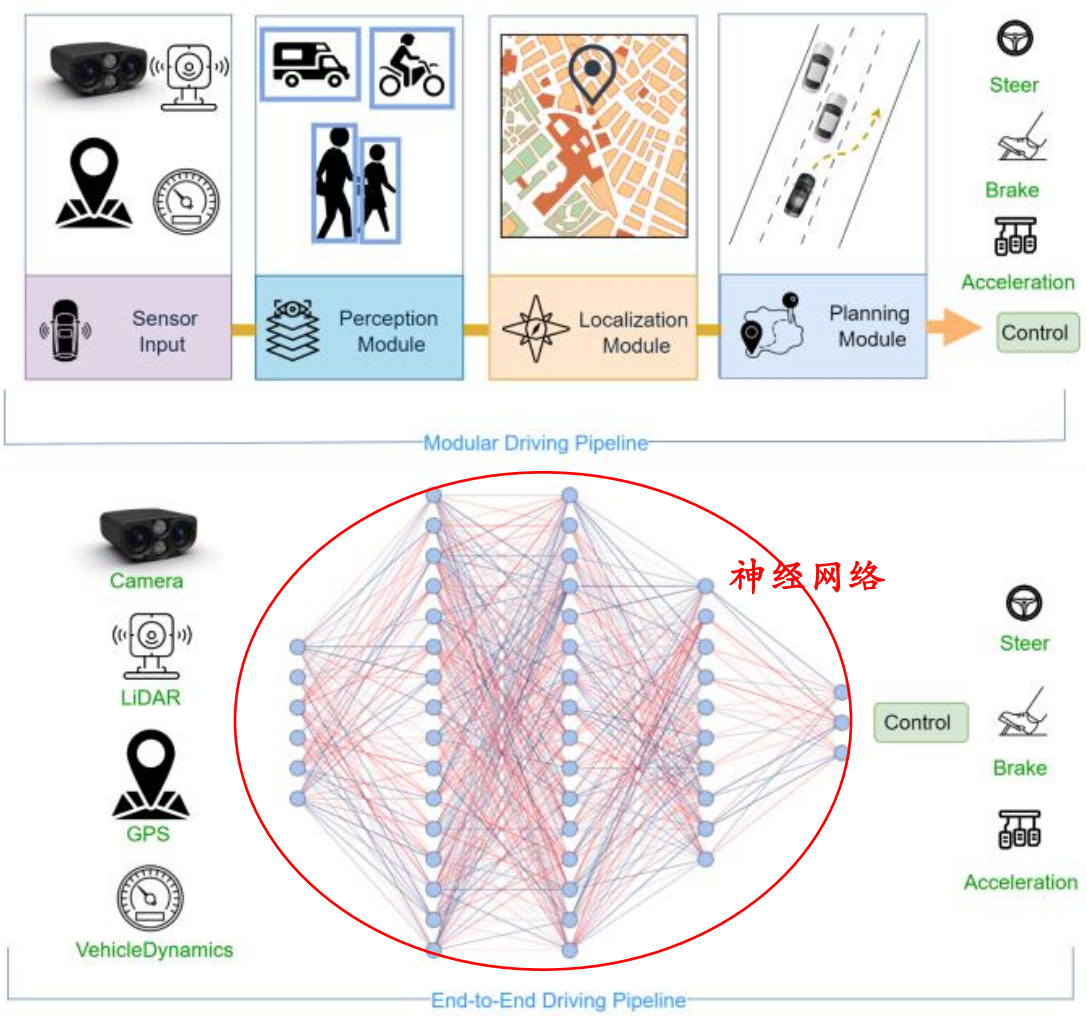
自动驾驶分类	原理	优点	缺点
模块化设计	将自动驾驶系统拆分为众多模块。每个模块可由基于规则的代码程序控制；也可以由训练好的机器学习或深度学习模型控制，每个模块的算法可以各不相同。	①安全、稳定、可靠； ②可解释性强，每个独立模块负责单独的子任务，便于问题回溯，易于调试等；	①系统庞大且复杂，涉及很多代码或算法； ②算力要求高，当越来越多的模块采用深度学习网络时，将引爆计算需求； ③多数情况下需要使用昂贵的激光雷达来确定障碍物的位置、并需要实时更新的高清大地图和其他辅助技术等； ④存在信息损失和误差问题；
端到端路线	将自动驾驶系统视为一个整体，而不是切分为模块，最后总只用一个模型来实现自动驾驶。例如。将传感器采集的信息直接送入深度学习神经网络，神经网络处理后直接输出自动驾驶的指令。	①成本小，降低对激光雷达、高精地图的依赖、减少中间环节的标注成本等； ②可借助数据的多样性获得不同场景下的泛用性； ③无需人工设计繁复的规则，深度学习神经网络通过训练数据就能学会驾驶； ④随着海量数据的自回归预训练，有望出现“智能涌现”；	①解释性差，当系统出现错误时，难以判断是哪个隐藏层或神经元的问题； ②闭环验证较难，缺少真实数据验证；

资料来源：清华大学智能产业研究院，西南证券整理

1.4.1 技术路径：大模型成就端到端自动驾驶，推动感知决策一体化

- 通过对比右侧的模块化和端到端两大技术**路径示意图**，我们更能直观地理解两者的区别：模块化方案由众多子模块组成，每个子模块对应特定的任务和功能；端到端则是输入感知信息并直接生成控制信号的单一路径。
- **从端到端自动驾驶技术路径来看，神经网络是关键，强化学习是重要方法。**神经网络结构受人脑启发，模仿生物神经元相互传递信号的方式，通过综合各种信号做出判断和反应。端到端自动驾驶主要学习方法为强化学习(RL/Reinforcement Learning)，即一种学习如何从状态映射到行为以使得获取的奖励最大的学习机制，在自动驾驶场景中，神经网络做出的驾驶决策由人类给予**奖励或处罚**等反馈，以此来不断优化驾驶行为。

模块化&端到端 自动驾驶技术路径示意图

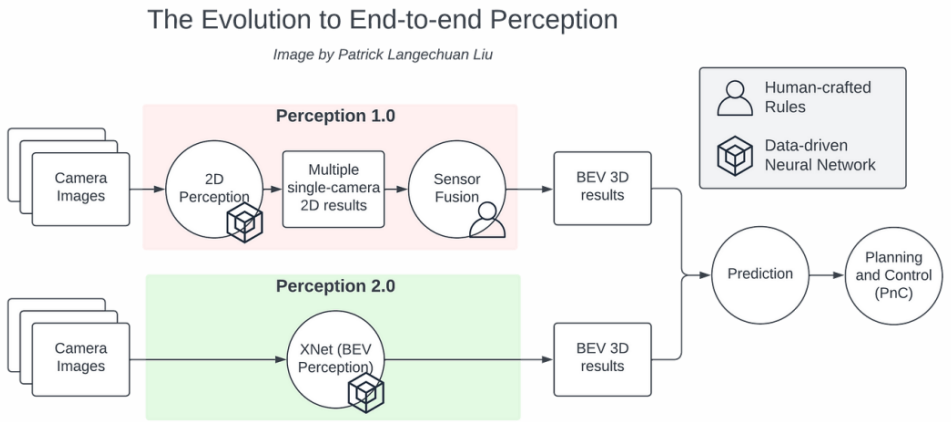


资料来源：《Recent Advancements in End-to-End Autonomous Driving using Deep Learning: A Survey》，西南证券整理

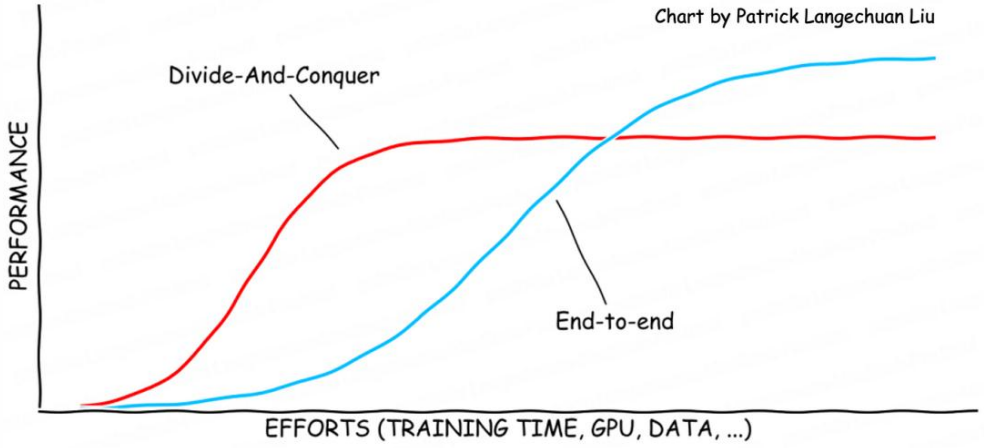
1.4.2 技术路径：车企率先聚焦端到端感知，BEV+Transformer成为主流

- 随着端到端技术的持续发展，其在自动驾驶系统中的**感知环节实现率先应用**，众多车企和算法公司基于Transformer架构做算法改进，BEV+Transformer逐渐成为主流解决方案。
- **BEV感知本质上是端到端感知解决方案。**在传统的自动驾驶堆栈中，2D图像被输入感知模块以生成2D结果，然后通过传感器融合方案将多个摄像机的2D结果转换为3D图像，以供系统进行预测和规划。在端到端感知中，BEV感知模型可以使车辆直接在BEV空间中感知环境，辅助自动驾驶。
- **端到端有望突破性能天花板，找到近似最优解。**对比分而治之和端到端两种解决办法，分而治之可以在有限的精力内快速实现性能的提升、并形成解决方案，但该方法容易陷入局部最优解，导致性能上限仅为80%。而端到端解决方案通过反复多次、集中优化一系列组件，从而不断突破性能天花板，直至实现完全的端到端解决方案，从而摆脱局部最优解的痛点，找到近似全局的最优解。

BEV感知本质上是端到端感知

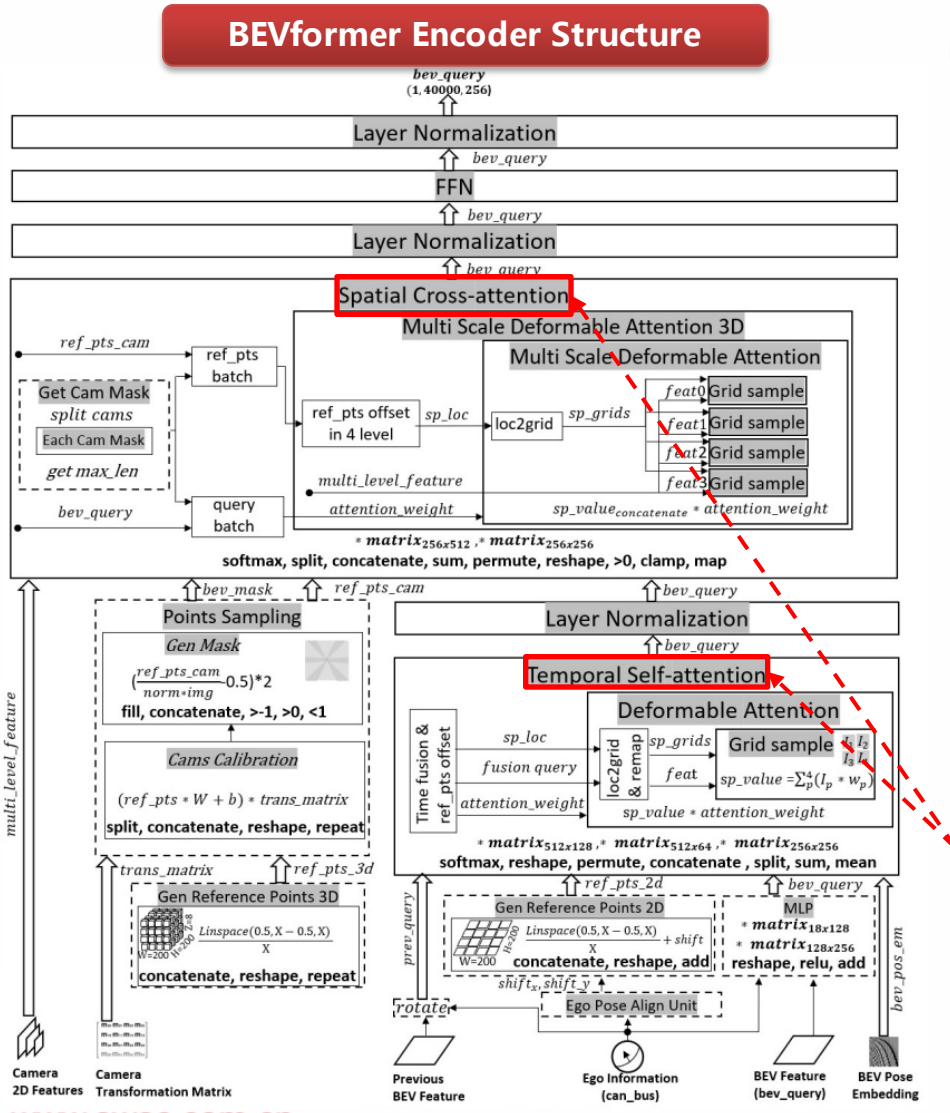


分而治之与端到端的性能增长曲线



资料来源：小鹏汽车，西南证券整理

1.4.2 技术路径：车企率先聚焦端到端感知，BEV+Transformer成为主流



- **Transformer架构在自动驾驶系统的感知环节中的运用优势：**
- ①Transformer在自然**语言处理**领域和计算机**视觉感知**领域均能发挥作用。
- ②Transformer在处理**大规模数据量场景**上具备优势，较神经网络可以更好地在海量图像数据中识别数据间的**关联关系**，更有利于**构建向量空间**。
- ③Transformer网络架构**引入注意力机制**，关注重要信息而非全部信息，在时间性方面具有更高的**并行计算效率**，在空间性能方面具有更强的**泛化能力**。

BEVformer编码器具有两种注意机制：

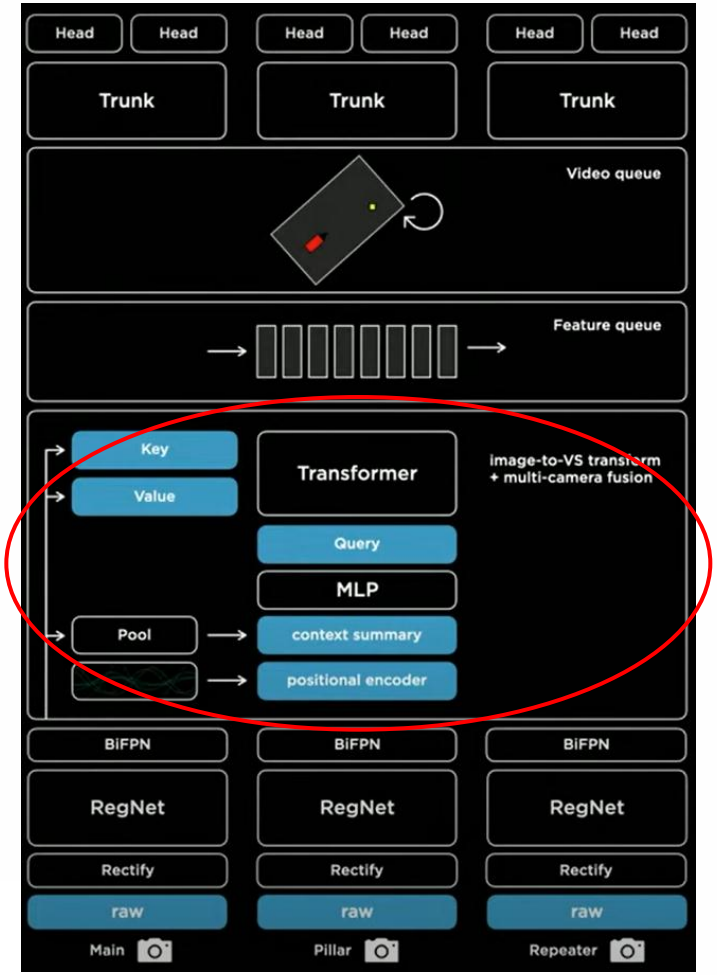
①时间自注意机制：通过自我信息校准对由previous BEV feature和current BEV feature初始化的bev query执行可变形注意(deformable attention)。

②空间交叉注意机制：从2D摄像头特征中提取BEV特征，且同样运用可变形注意机制，采用多摄像头query，增加两大模块，一是摄像头掩模模块，可生成BEV空间中的每个摄像头掩模，另一个是多级偏移模块，可获得4个级别的参考点偏移。

资料来源：《Transformer-based models and hardware acceleration analysis in autonomous driving: A survey》，西南证券整理

1.4.2 技术路径：车企率先聚焦端到端感知，BEV+Transformer成为主流

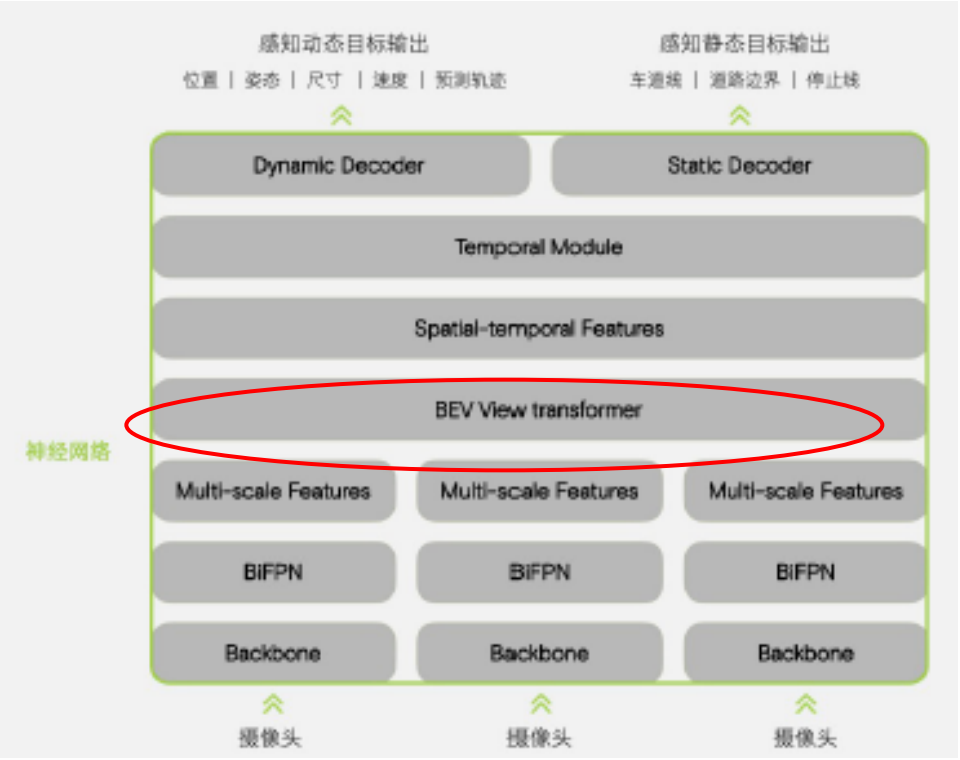
特斯拉感知架构HydraNet



资料来源：Tesla AI day，西南证券整理

- **采用端到端感知方案的代表企业包括特斯拉、小鹏汽车等。**2021年特斯拉于AI Day首次在算法层面引入Transformer，与此同时，小鹏汽车等国内车企也积极引入Transformer架构，改进自身算法，并在更短时间内完成了对架构的重写。

小鹏汽车感知架构Xnet



资料来源：小鹏汽车官网，西南证券整理

目 录

2.1 硬件端：全栈自研HW3.0，底层硬件继续向更高级别迭代

- **HW1.0向HW3.0快速迭代，硬件性能有望持续升级。1）HW1.0：**2014年10月，特斯拉基于Mobileye芯片Mobileye EyeQ3发布第一代硬件Hardware1.0。**2）HW2.0：**2016年10月，特斯拉推出HW2.0，芯片由英伟达提供，并配置8个摄像头+12个远程超声波雷达+1个前置毫米波雷达，在功能上实现辅助驾驶，且该配置延续至Hardware3.0。**3）HW3.0：**2019年4月，特斯拉发布Hardware3.0系统，采用全栈自研FSD芯片，单个芯片算力达72TOPS，远高于当时市面上的自动驾驶芯片，算力实现大幅提升，在功能上可识别更多目标。**4）目前，特斯拉正处于由HW3.0向HW4.0更高级别硬件的迭代阶段，未来有望支持4D毫米波雷达等更多传感器和摄像头的接入，使GPU集成化更高、模块更轻薄，FSD芯片内核数量有望持续增多，进一步提升性能等。**

特斯拉自动驾驶硬件迭代历程

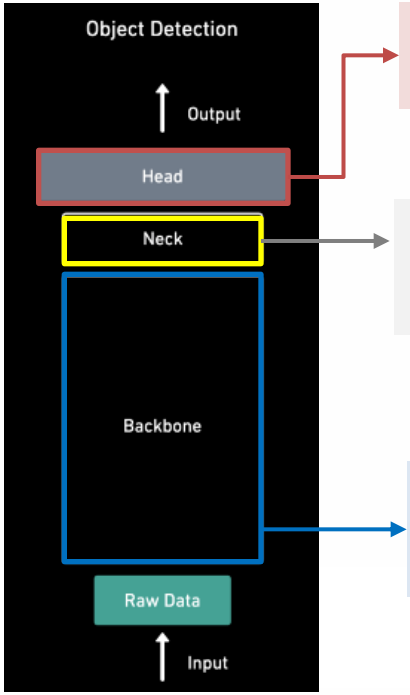
	HW1.0	HW2.0	HW2.5	HW3.0
前置摄像头	1个	1*Camera 35° /1*Camera 50° /1*Camera 120°		
侧面相机	0	2*Camera 90°		
侧面后置摄像头	0	2*Camera 60°		
毫米波雷达	1*Radar 160m		1*Radar 170m	
超声波雷达	12*Lidar 5m	12*Lidar 8m		
核心处理器	1*Mobileye EyeQ3	1*NVIDIA Parker SoC + 1*NVIDIA Pascal GPU + 1*英飞凌 TriCore MCU	2*NVIDIA Parker SoC + 1*NVIDIA Pascal GPU + 1*英飞凌 TriCore MCU	2*FSD 芯片
ROM	256兆字节	6GB	8GB	2*8GB
Flash	/	/	/	2*4GB
处理能力	1倍	40倍	40倍带冗余	420倍带冗余
每秒处理帧数	36	110	110	2300
估计功率	25W	250W（闲置40W）	300W	220W

资料来源：焉知汽车，西南证券整理

2.2 算法端：神经网络为基，迈向大模型时代

- **1）2016-2018年：通用网络结构阶段。**在自动驾驶行业发展初期，业内车企在自动驾驶的目标检测上一般采用通用网络结构(Input→backbone→neck→head→Output)，该结构中仅有一个head，是单一的目标检测，而驾驶场景通常面临多项任务，如车道线/人物/信号灯检测等，因此单一检测难以满足现实需求。**2）2018-2019年：多任务学习神经网络阶段。**为解决单一检测的痛点、能够完成多头任务，特斯拉构建出多任务学习神经网络架构**HydraNet**，并使用特征提取网络BiFPN，实现多特征共享和多任务处理，提升算法效率。**3）2020年至今：大模型时代。**特斯拉引入特征级融合和BEV+Transformer，特征级融合使原始数据的融合效果提升，BEV使摄像头拍摄的2D视角转变为3D表达，Transformer通过适应不同形式的输入使得BEV在自动驾驶领域得以实现。

目标检测的通用网络结构

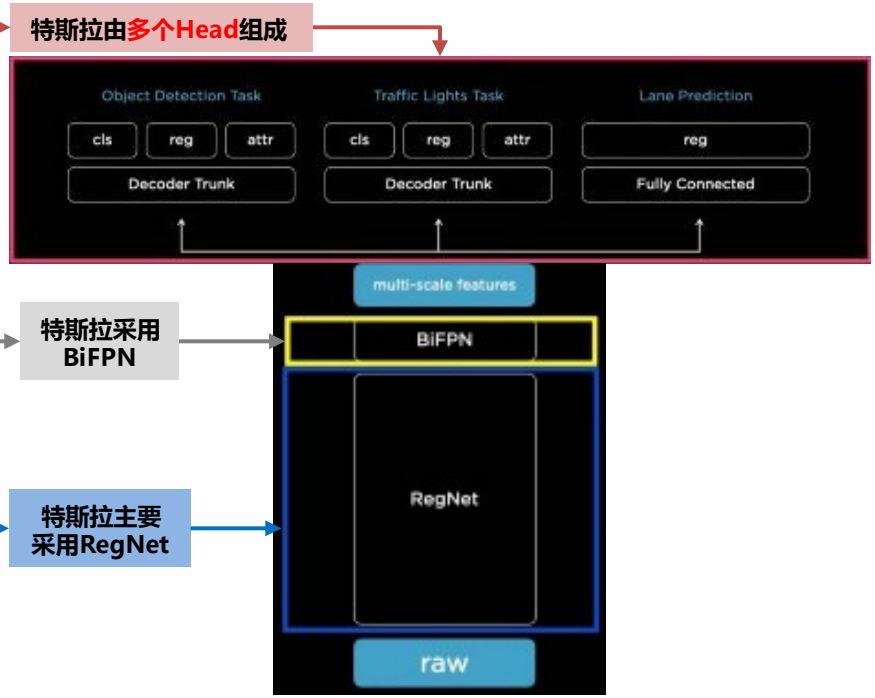


Head：进行具体的下游任务，如物体检测、交通信号和车道识别等

Neck：提取更复杂的特征，例如使用特征金字塔网络FPN、BiFPN等提取不同尺度的特征

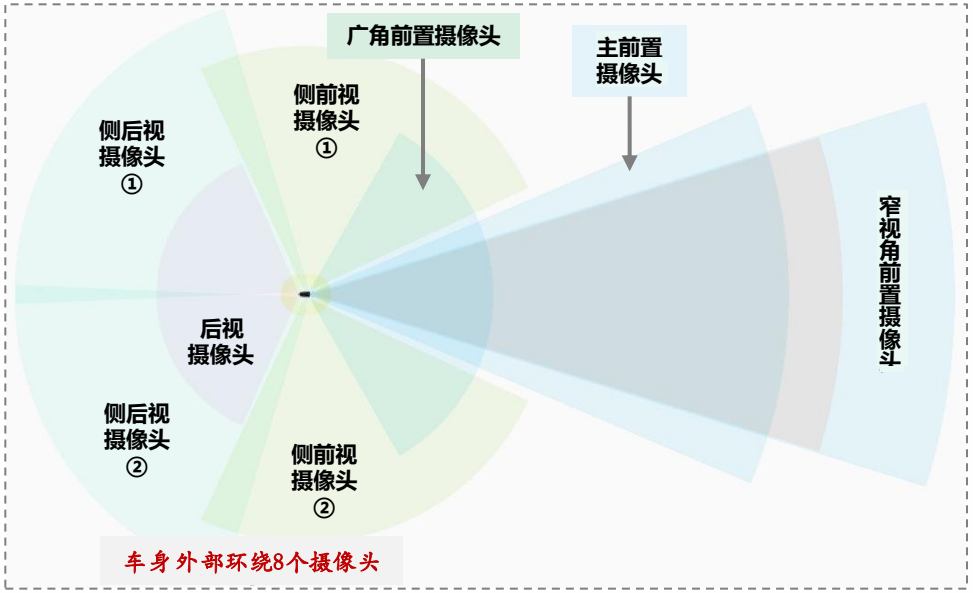
Backbone：提取图像特征，网络结构通常包括AlexNet、ResNet、VGGnet、Densenet等

特斯拉多任务学习神经网络架构HydraNet

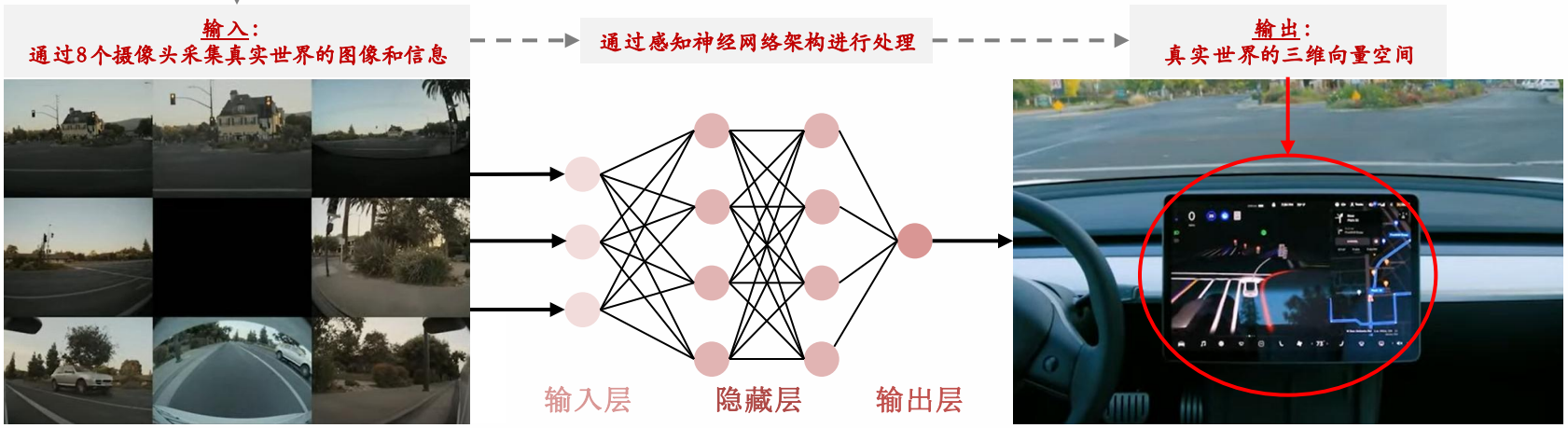


资料来源：汽车之心，西南证券整理

2.2.1 感知算法：采用端到端感知架构，构建三维向量空间



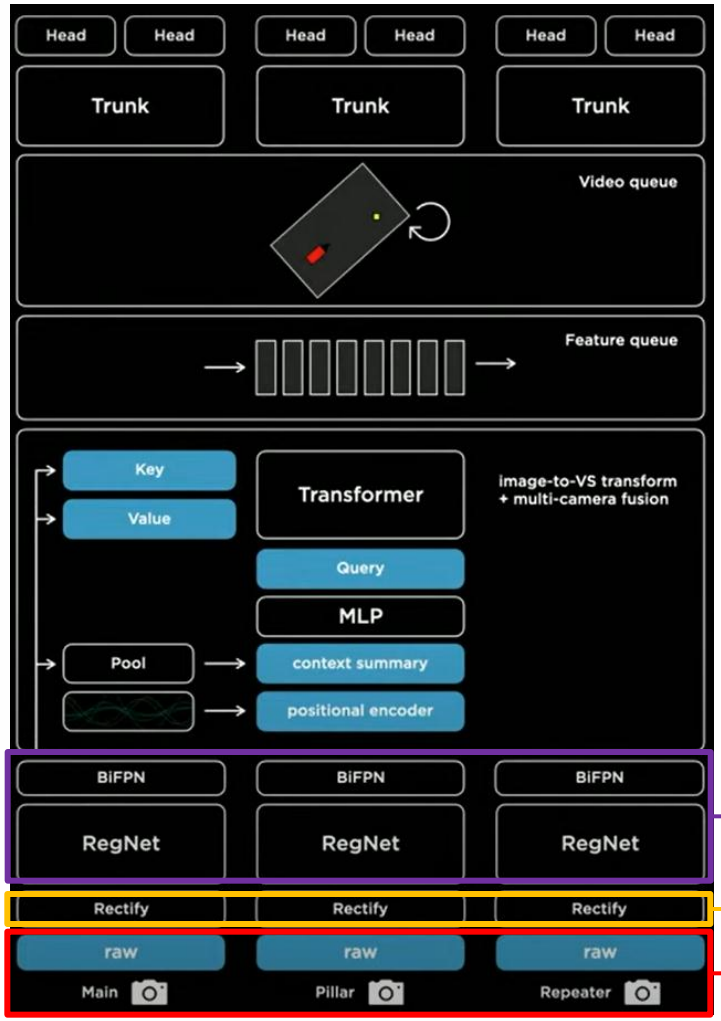
- **端到端感知：输入多相机图像，输出三维向量空间。**特斯拉车身外部环绕8个外部摄像头，摄像头对车身周围环境的图像数据和信息进行采集，再通过感知神经网络进行处理，系统通过深度学习模型进行自我培训，从而达到全范围认知路况、增进系统控制精度的目的，构建真实世界的三维向量空间，其中包含汽车、行人等动态交通参与物，道路线、交通标识、红绿灯、建筑物等静态环境物，以及各元素的坐标位置、方向角、距离、速度、加速度等属性参数。



资料来源：Tesla AI Day，西南证券整理

2.2.1 感知-数据校准层：构建标准化数据，实现多机位融合

特斯拉视觉感知网络架构



- **数据输入层/Input：多机位**，且每个摄像头每秒输入36帧12位1280×960的高清图像。
- **数据校准层/Rectify：**摄像头外参差异会导致采集的数据出现偏差，因此在感知框架中加入“虚拟标准相机”，通过去畸变、旋转等处理，将图像数据统一映射到同一虚拟标准摄像头坐标中，保证数据一致性。
- **RegNet网络和BiFPN：**通过RegNet网络和BiFPN进行特征提取，感知不同尺度的目标，采用多尺度特征融合方法，获得160×120×64、80×60×128、40×30×256、20×15×512四个尺度的特征图。

RegNet网络和BiFPN

数据校准层/Rectify

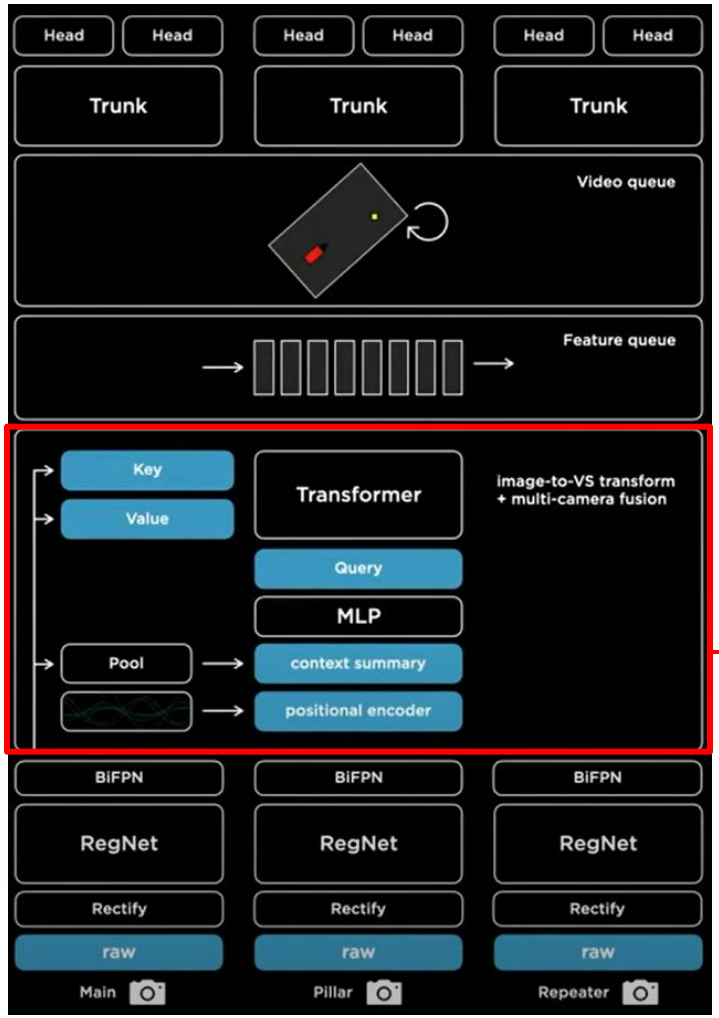
数据输入层/Input

构建标准化数据，实现多机位融合

资料来源：Tesla AI Day，西南证券整理

2.2.1 感知-空间理解层：2D数据实现3D变换，自动驾驶接近现实

特斯拉视觉感知网络架构



- **特斯拉通过引入Transformer，使自动驾驶的思维方式接近真实世界。**特斯拉在BEV空间层对图像进行特征初始化，再通过多层Transformer与2D图像特征进行交互融合，最终得到BEV特征，实现BEV视角的转换。
- **特斯拉运用Transformer的多头注意力机制将每个摄像头的图像转换为key和value，然后训练模型以查表的方式自行检索需要的特征用于预测。**具体来看，Key和Value由多尺度特征空间经过多层感知机网络(MLP)训练得到。而通过对特征空间进行池化处理得到全局描述向量(context summary)，同时对输出的BEV空间各栅格进行位置编码(positional encoder)，合成描述向量和位置编码后再通过一层MLP可以得到Query。

空间理解层/Transformer：

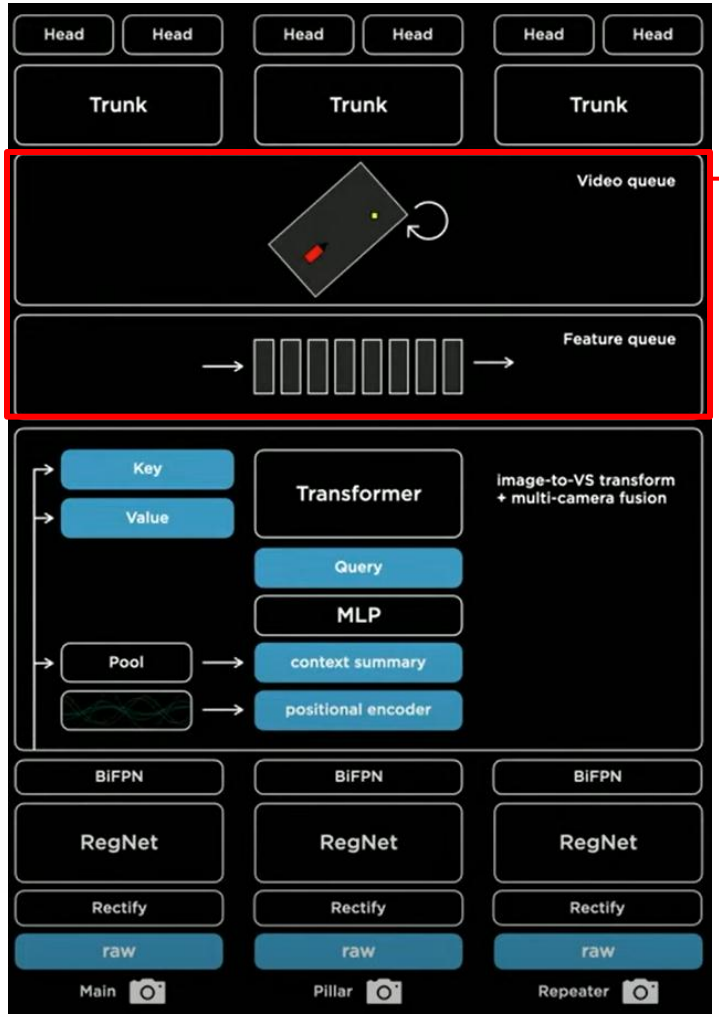
二维数据实现三维变换，自动驾驶接近现实世界

- 特斯拉通过运用Transformer，可以将地面坡度、曲率等几何形状的变化情况内化进神经网络的训练参数中，实现对物体深度信息准确感知和预测，例如，车道线更准确、清晰，目标检测结果更稳定，不再有重影等。

资料来源：Tesla AI Day，西南证券整理

2.2.1 感知-短时记忆层：充分考虑时序信息，模拟司机短期记忆

特斯拉视觉感知网络架构



- **特征队列模块(Feature queue module)**用来**缓存**时序特征，**视频模块(Video module)**用来**融合**时序上的信息，经模块处理后的特征融合了时序上的多相机特征，最终在Heads中进行解码并实现输出。

短时记忆层：

充分考虑时序信息，模拟司机短时记忆

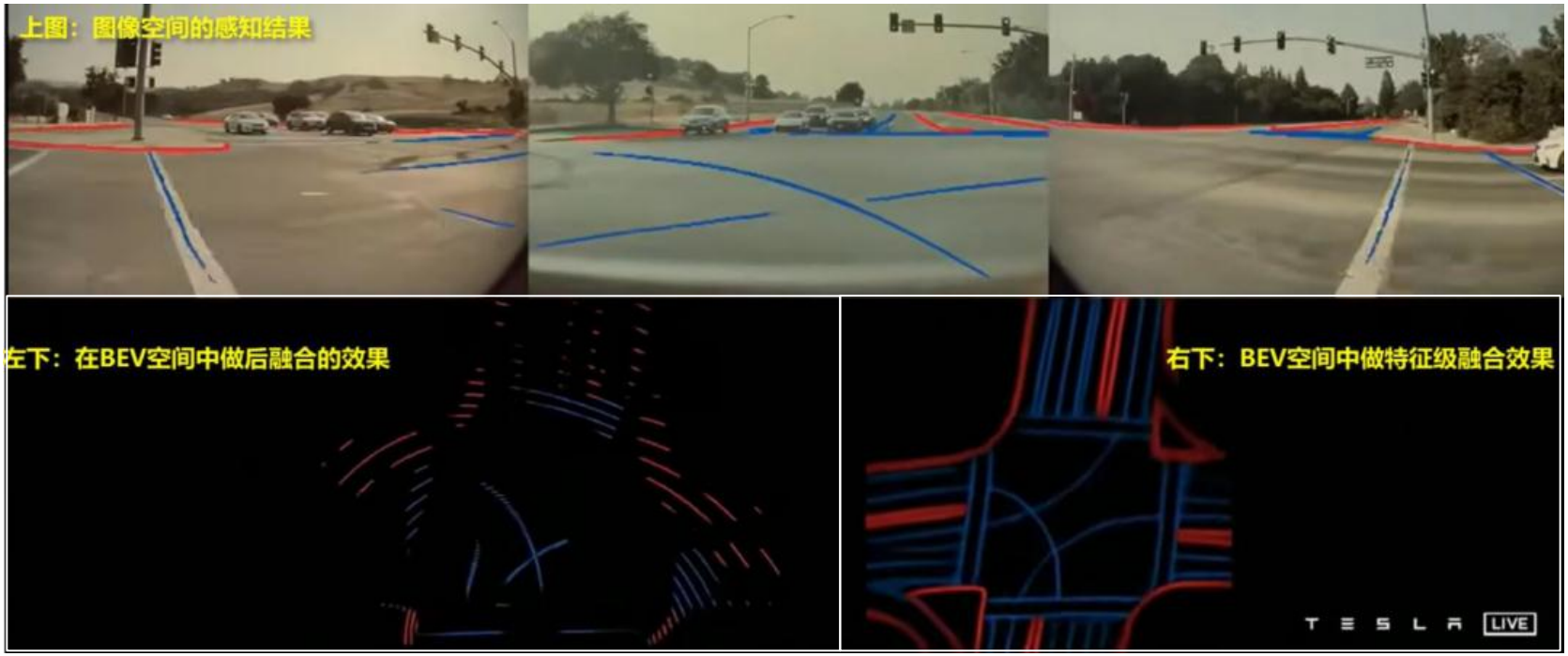
- **特征队列模块**：可分为时序特征队列和空间特征队列。**1）时序特征队列**：每过27ms将一个特征加入队列。时序特征队列可以稳定感知结果的输出，如运动过程中发生的目标遮挡，模型可以找到目标被遮挡前的特征来预测感知结果。**2）空间特征队列**：每前进1m将一个特征加入队列，用于等红绿灯一类需要长时间静止等待的状态，在该状态下，一段时间之前的在时序特征队列中的特征会出队而丢失，因此需要用空间特征队列记住一段距离之前路面的箭头或路边的标牌等交通标志信息。
- **视频模块**：特斯拉使用**RNN结构**来作为视频模块，命名为空间RNN模块(Spatial RNN Module)。**添加视频模块能够提升感知系统对于时序遮挡的鲁棒性、对于距离和目标移动速度估计的准确性。**
- **短时记忆层**使得自动驾驶感知网络拥有类似于司机的短时记忆，可以对**当前时刻的场景做出判断**，并根据一段时间内的数据特征**推演出目前场景下的可能结果。**

资料来源：Tesla AI Day，西南证券整理

2.2.1 感知：特征级融合取代后融合，降低算力消耗和复杂度

- **感知信息采用特征级融合，拟合效果显著优于后融合。**特斯拉起初采用后融合方案，但在后融合方案下，置信度较低的信息容易被忽略，原始数据也容易丢失，从而会导致信息失真、决策失误等问题。而特征级融合可以避免不同的摄像头对同一特征进行识别，因此能够更好地解决后融合信息失真的问题。根据特斯拉AI Day展示的效果图来看，在BEV空间中做特征级融合的效果要远远好于后融合，同时能够避免前融合方案下的巨大算力消耗、以及后融合方案下的复杂度难题。

特斯拉在BEV空间中做后融合和特征级融合的效果对比



资料来源：Tesla AI Day，西南证券整理

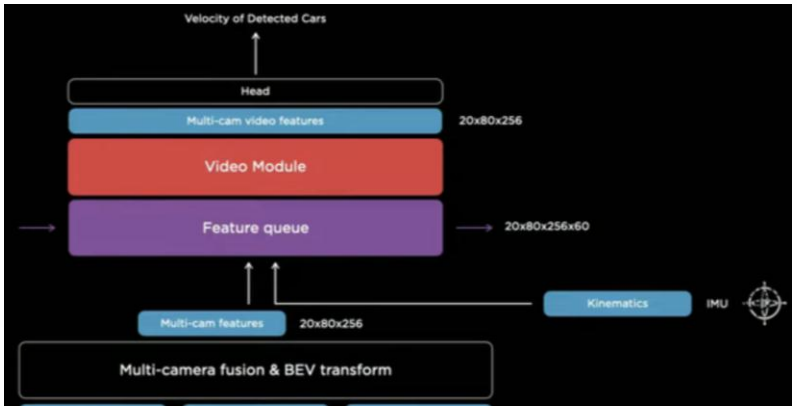
2.2.1 感知：BEV+Transformer提供全局视角，应对长尾问题

- **时序信息叠加占用网络，减轻长尾难题。**目前，自动驾驶技术已成功运用于大多数道路场景，但在部分长尾场景和极端情况中，自动驾驶算法的泛化能力仍然难以支持，而BEV+transformer技术通过提供全局视角和增强推理能力来优化自动驾驶系统的感知能力，进一步提高系统的可靠性和安全性。在特斯拉的BEV+Transformer架构中，通过增加时序信息、引入占用网络，**算法的泛化能力进一步提升，从而能够应对更多的长尾场景**(Corner Case)。

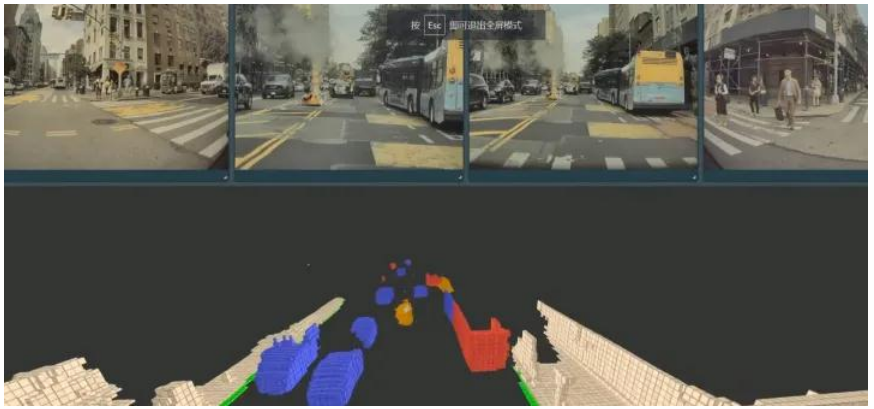
- **增加时序信息：**例如，在自动驾驶过程中，如果有行人正在过马路，该过程却被障碍物遮挡，如果汽车仅有瞬时感知能力，那么汽车在感知时刻则无法感知到可能存在行人被遮挡的情况，从而产生安全风险。但人类司机可以根据过往经验以及有关行人过马路的历史记忆，意识到行人被遮挡的风险，从而选择减速避让。因此，特斯拉在自动驾驶感知网络架构中**引入时空序列特征层，使用视频片段而非静止图像来训练神经网络，通过某一时间段的数据特征，推演当前场景下可能性最大的结果。**

- **引入占用网络：**占用网络即“不考虑某一物体究竟是什么，只考虑体素是否被占用”，体素的作用相当于激光雷达点阵的作用，使非典型但却存在的事物能够直接表示出来，从而**增加算法的泛化能力和对现实世界的精确认知。**

特斯拉在感知网络引入时空序列特征层



特斯拉占用网络在行车场景中的使用效果



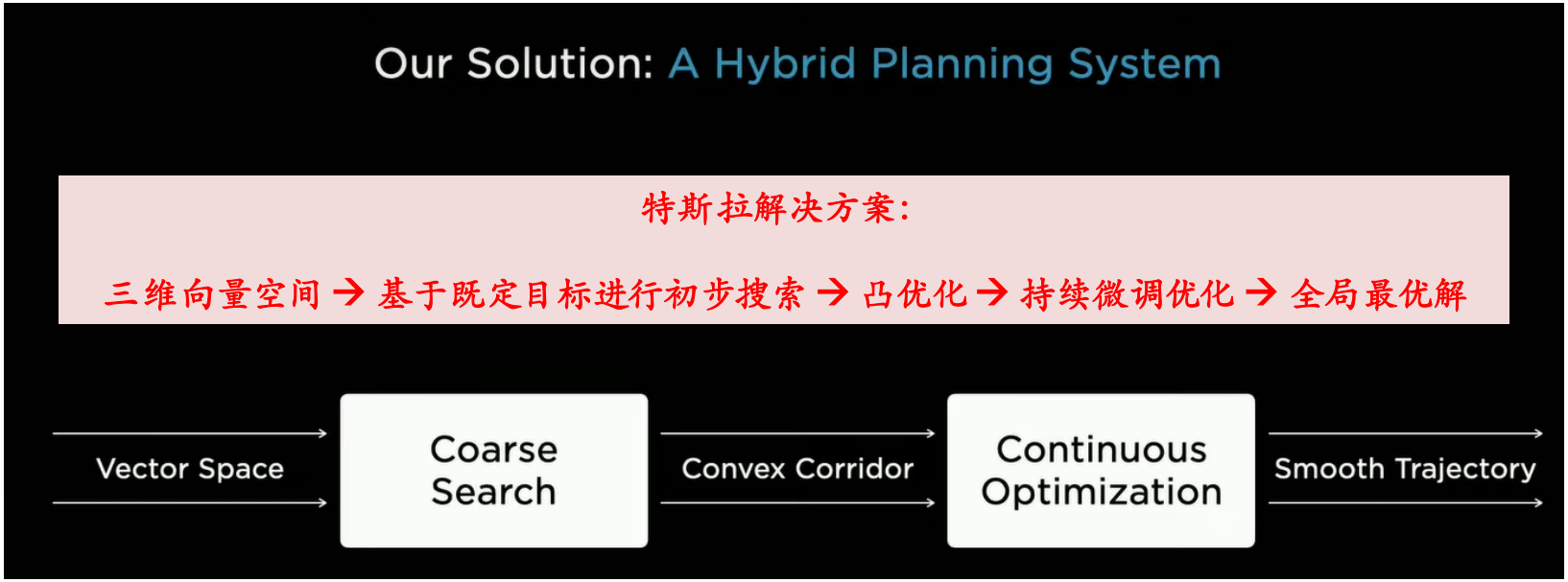
注：蓝色表示运动的体素，红色表示静止的体素

资料来源：Tesla AI Day，西南证券整理

2.2.2 规控算法：推出自有解决方案，寻找规控最优解

- **规控两大难点：1）非凸性(Non-Convex)：**通常一个问题有众多解决方案，难以得出全局最优解，此外，将离散搜索转化为连续的函数优化，使用梯度下降算法容易陷入局部最优，从而无法快速做出准确决策。**2）多维的参数量/参数空间分布广：**车辆行驶包括众多参数，快速决策的要求会增加搜索和计算的复杂程度。
- **特斯拉解决方案：1）对于仅有唯一解的问题：**直接生成明确的规控方案。**2）对于有多个可选方案的复杂问题：**在由感知获得的三维向量空间中，基于既定目标进行初步粗略搜索，寻找初步路径，然后根据安全性、舒适性、效率性等指标，围绕初步路径进行优化，再融入成本函数、人工干预数据、仿真模拟数据等，在障碍物间距、加速度等参数上继续微调，最终得到最优规控方案，最终生成控制指令，由执行模块接受指令，实现自动驾驶。

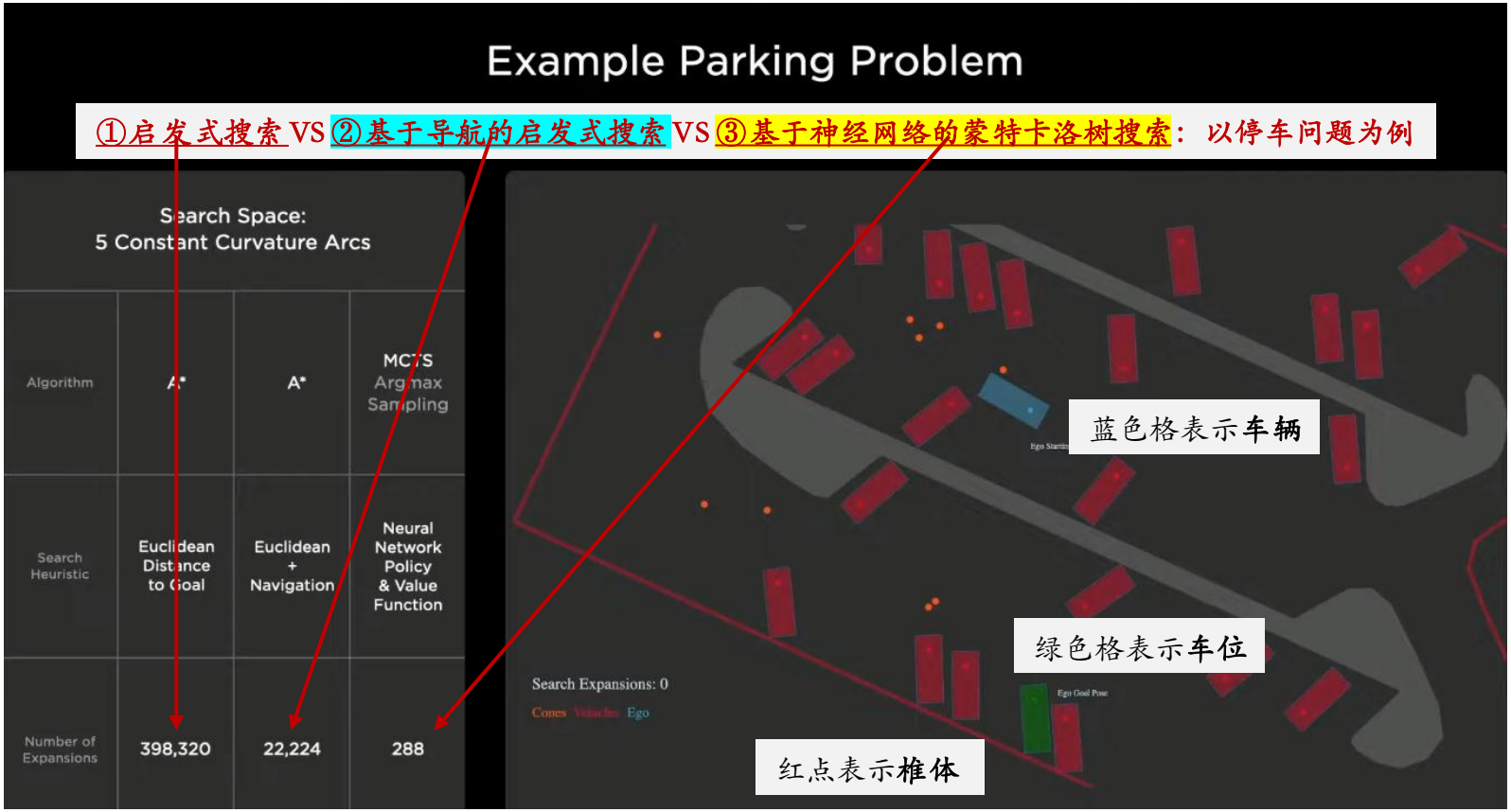
特斯拉视觉规控解决方案



资料来源：Tesla AI Day，西南证券整理

2.2.2 规控算法：引入蒙特卡洛树搜索，完成高效求解

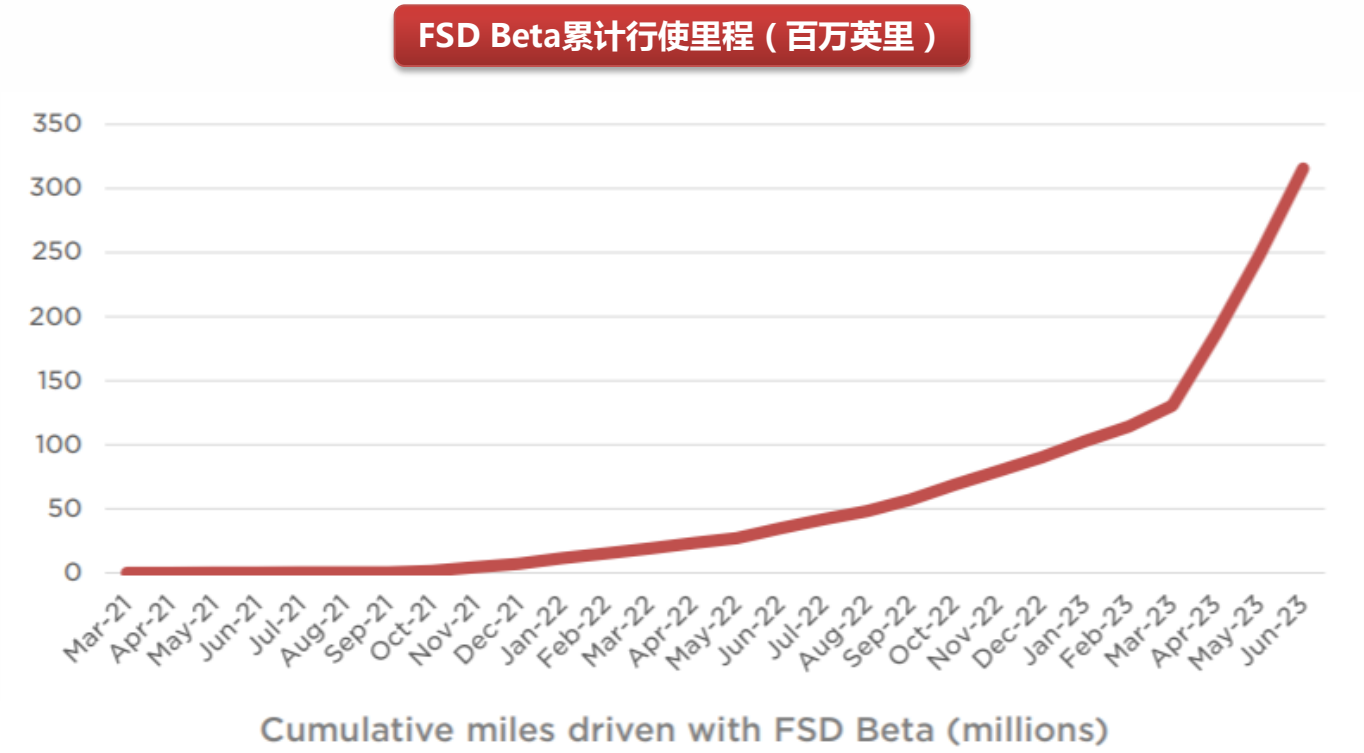
- **引入蒙特卡洛树搜索，完成高效求解。1）启发式搜索：**采用A-Star算法，搜索所有可能路径，直至找到解决办法，以停车问题为例，需要近40万次搜索。**2）基于导航的启发式搜索：**在已知停车场地图的情况下按照导航寻找停车位，寻找效率相比启发式搜索提高10+倍，但椎体等未知目标会导致搜索效率降低。**3）基于神经网络的蒙特卡洛树搜索：**通过神经网络预测各个节点概率或状态，将节点放于蒙特卡洛树中进行搜索，有助于大幅减少搜索空间、有效提高决策实时性，在停车问题中仅需288次搜索即可完成求解，效率相比A-Star算法提高千倍。



资料来源：Tesla AI Day，西南证券整理

2.3 数据端：车队逐渐壮大，里程数日益增长，构建数据护城河

- **FSD里程数实现迅速增长：**根据特斯拉2023年上半年业绩会，**特斯拉FSD在里程数上取得新进展，FSD Beta累计行使里程已超过3亿英里**，仅23Q2单季度提升约1亿英里。
- **自身数据库反哺模型性能：**特斯拉车队规模逐渐壮大、车辆累计行使里程日益增长，有助于特斯拉构建自身的自动驾驶数据仓库，形成数据壁垒，为大模型的训练和优化提供更多的优质数据，反哺算法性能。

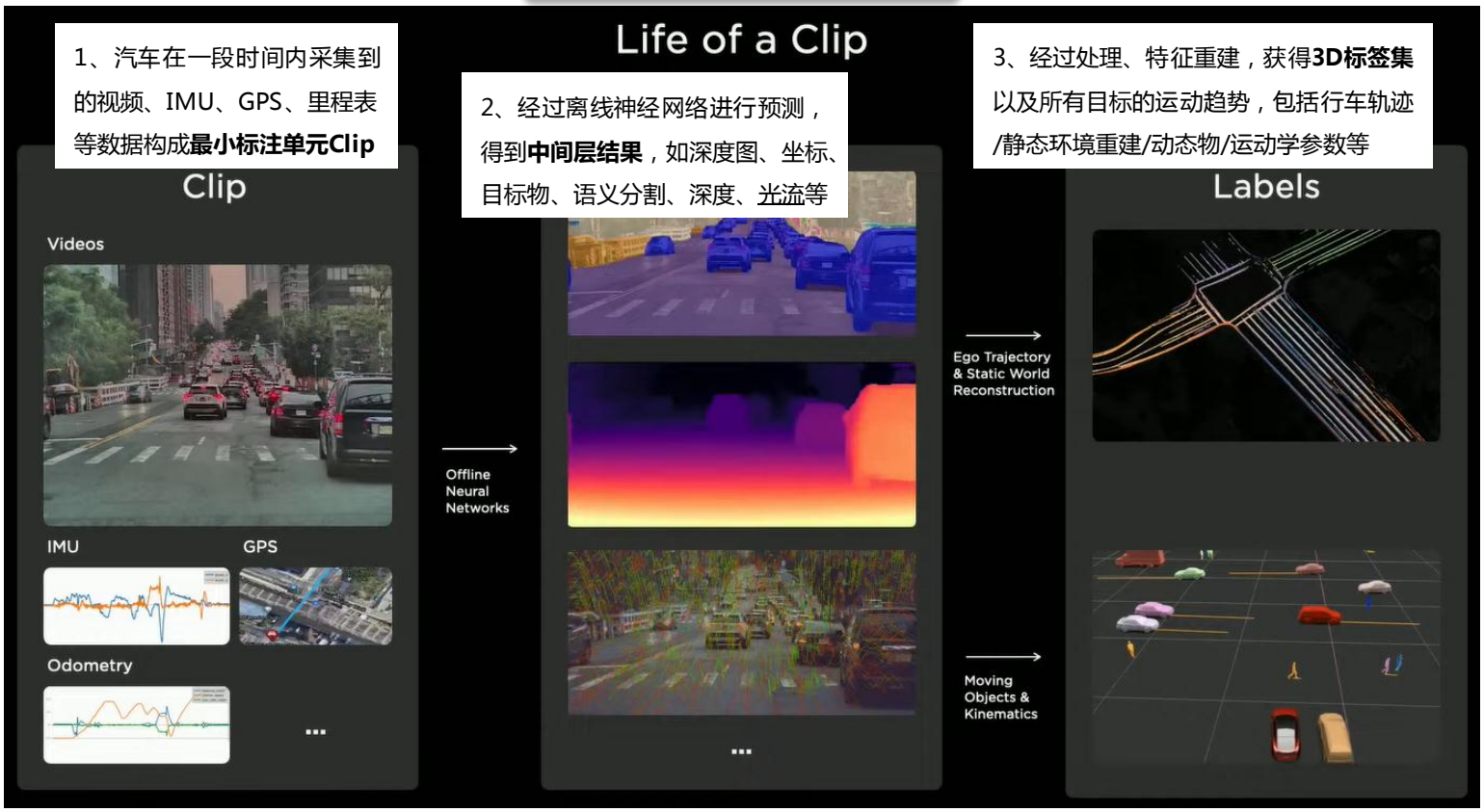


资料来源：公司公告，西南证券整理

2.3 数据端：数据标注从人工标注到自动标注

- **自建数据标注团队，保证标注质量及效率。**2018年，特斯拉与第三方公司合作，采用人工标注，该方式标注效率低、且沟通成本高。而后为提升标注效率和质量，特斯拉**自建标注团队**，人员规模近千人。此后，随着自动驾驶数据持续增长，所需标注人员的规模进一步扩大，使得人力成本快速增长，使得2020年特斯拉开始研发并使用数据自动标注系统，通过大量数据训练大模型，再用大模型训练车端小模型。

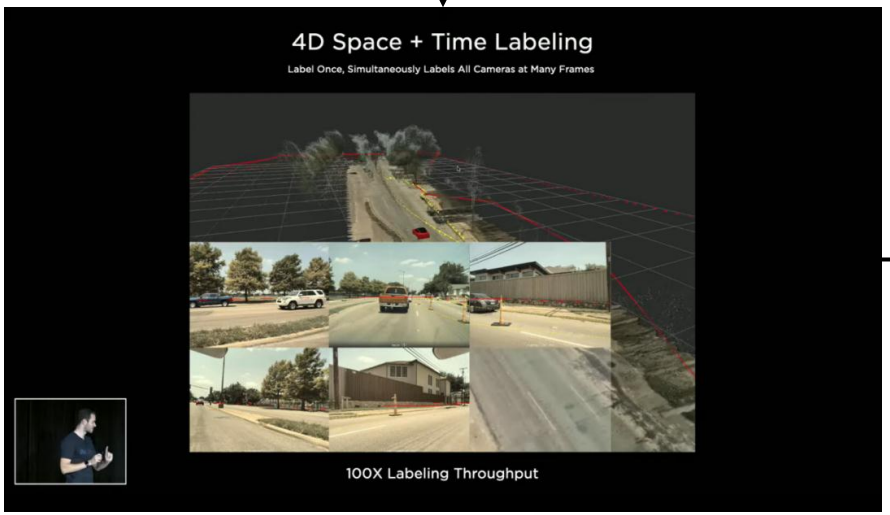
特斯拉数据标注流程



资料来源：Tesla AI Day，西南证券整理

2.3 数据端：数据标注从2D到3D、4D

数据标注从2D到3D、4D

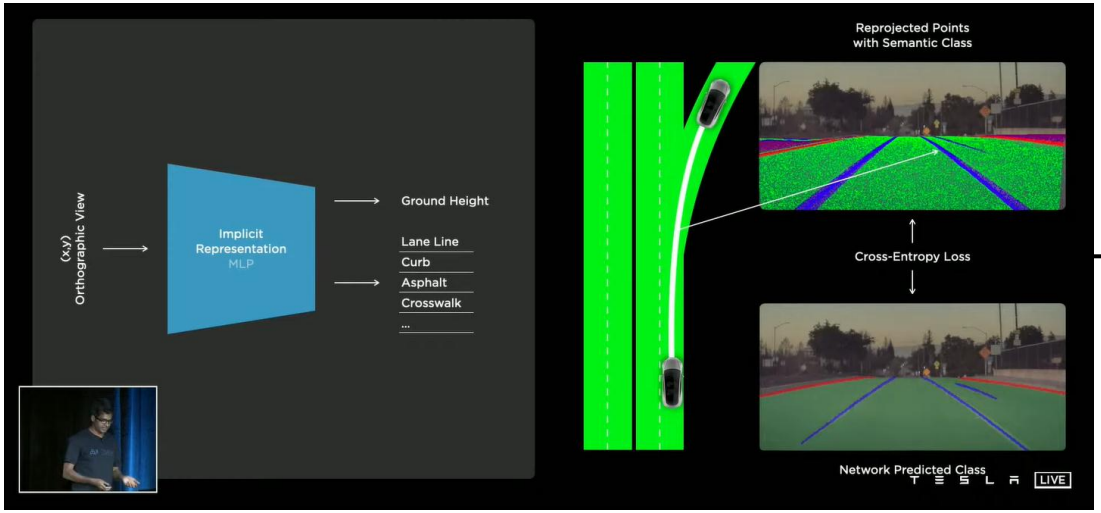


- **数据标注从2D到3D、4D：**
- 通过引入BEV视角，自动驾驶实现2D图像向3D车身自坐标系转变，但在未引入时序信息时，BEV仍然是对瞬时的图像片段进行感知，缺乏时空记忆力，汽车只能根据当前时刻感知到的信息进行判断。
- 特斯拉感知网络架构引入时空序列特征层，使用视频片段，而非用图像训练神经网络，增加短时记忆能力，并于2022年对BEV进行升级，引入占用网络，推动数据标注向4D升级。

3D点云重建图：
3D场景中的标签可以和2D场景中的标签相互转换；3D、4D数据可通过目标移动、方向转换，获得不同角度、视野的2D图像。

资料来源：Tesla AI Day，西南证券整理

2.3 数据端：语义重构，展现完整道路情况



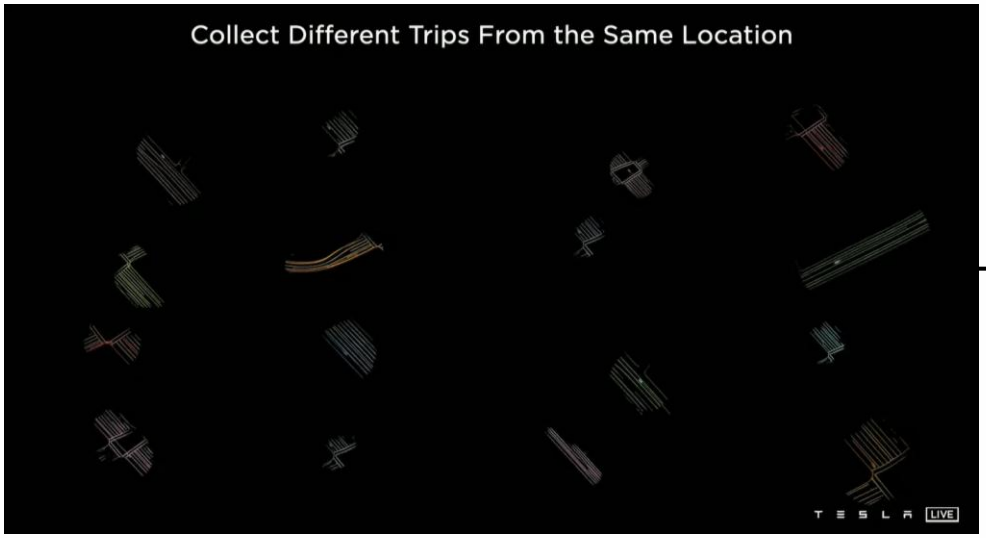
- 图像重构将2D图像的像素映射到对应的语义信息中，即一个2D的像素对应一个向量空间中的像素语义。

车辆在路面上行驶时，通过神经网络的隐含映射，每个像素都有对应的语义信息。

车辆在行驶时对拍摄的路面进行语义重构，可以绘制出整个道路的情况。

资料来源：Tesla AI Day，西南证券整理

2.3 数据端：语义重构，展现完整道路情况



- 不同车辆经过相同位置时，拍摄到的信息不同，通过多个车辆拍摄到的结果进行融合，可获得最新路况，而未经过的车辆也可以根据其他车辆走过的信息进行预判。

多个车辆经过同一位置时绘制的不同道路语义图。

- 多目标/车辆结果融合可以提高路况信息的准确性、减少数据误差、噪声等。

通过融合技术重构路面信息、3D点云等信息。

资料来源：Tesla AI Day，西南证券整理

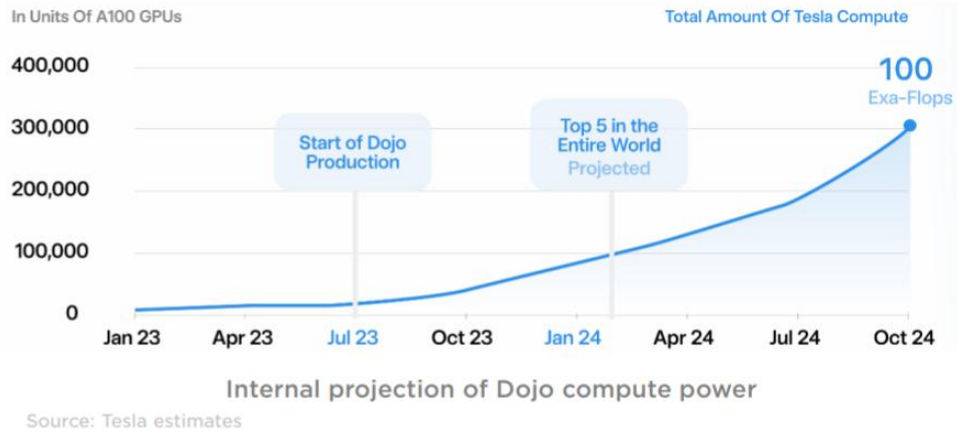
2.4 算力端：自研大规模集群超算平台，Dojo有望提供强算力

- **特斯拉自研超级计算平台Dojo——基于超大计算集群设计。**Dojo架构由特斯拉完全定制，涵盖计算、网络、输入/输出（I/O）芯片、指令集架构、电源传输、冷却等，具备高可扩展性和分布式系统。Dojo具备超高集成度，并非根据小系统拓展而来，旨在高效地处理海量视频数据、进行定制的神经网络训练。Dojo于2021年首届特斯拉AI Day上面市，当时仅有第一批芯片和训练块，尚未构建起完整的Dojo机柜和集群（Exapod）；2022年AI Day，Dojo取得新进展，并通过后续的持续部署与规划，搭建起大规模算力集群，推动大模型训练。
- **算力规划明确，Dojo正式投产。**1）2023年7月，Dojo进入投产阶段，拉开特斯拉算力集群快速建设阶段的帷幕；2）预期2024年2月，特斯拉的算力规模进入全球前五；3）预期2024年10月，特斯拉的算力总规模达到100EFlops，相当于30万块A100GPU的算力总和。

特斯拉Dojo的构成

资料来源：Tesla AI Day，西南证券整理

特斯拉Dojo算力规划

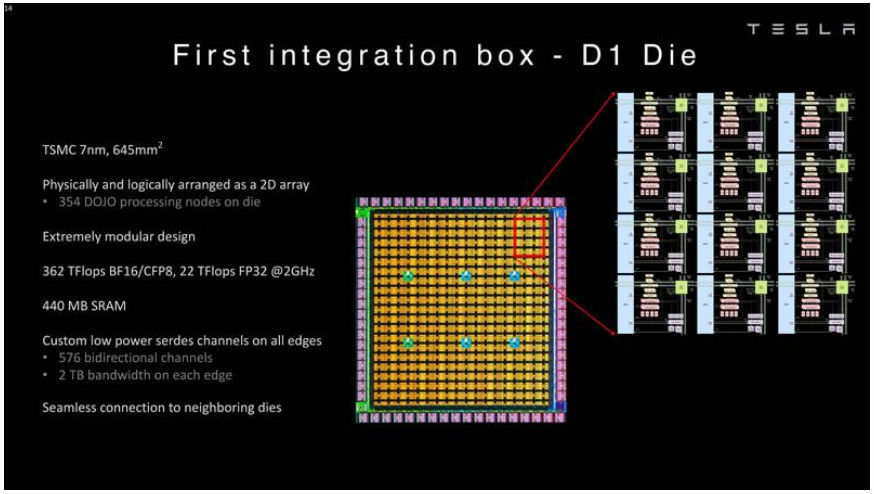


资料来源：公司公告，西南证券整理

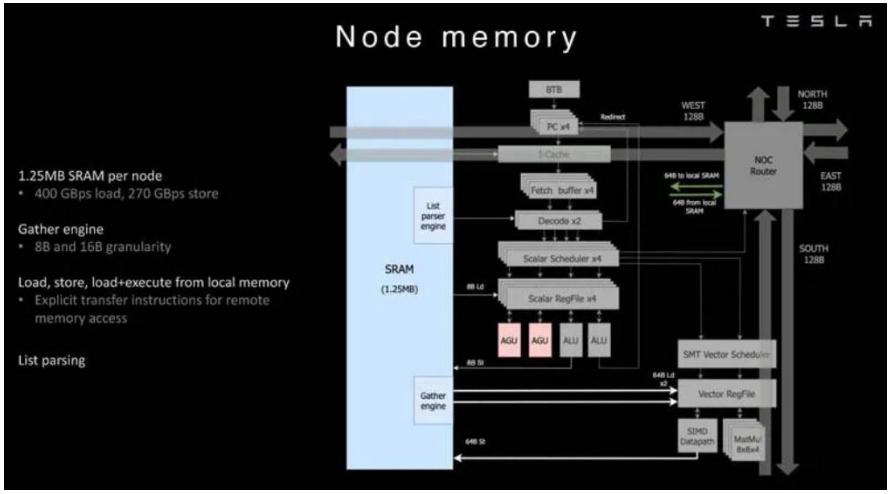
2.4 算力端：特斯拉Dojo——自研的D1芯片

- **Dojo D1性能**：Dojo的计算核心采用特斯拉自研的D1芯片，D1芯片使用台积电7nm工艺，拥有500亿个晶体管，芯片面积为645mm²，BF16、CFP8算力可达362TFlops，FP32算力可达22.6TFlops，TDP为400W。而英伟达A100芯片同样采用台积电7nm工艺，拥有542亿晶体管，芯片面积826mm²，FP32峰值算力为19.5TFlops。
- **Dojo D1架构**：D1芯片由18×20颗核心构成，出于良率和稳定性考虑，每个D1芯片有354颗核心（Node）可用。从每颗核心的微架构来看，**D1 Node采用存算一体架构（近存计算）**，带有向量计算/矩阵计算能力的处理器，具有完整的取指、译码、执行部件，处理器运行在2GHz，具有4个8x8x4矩阵乘法计算单元。同时，每个内核拥有一个1.25MB的SRAM作为主存（非缓存），能以400GB/S的速度进行加载，并以270GB/S存储。可以看出，**每个D1核心都是一个完整的带矩阵计算能力的CPU**，且特斯拉对其进行高计算密度的优化，其计算灵活性远超众核架构GPU，但同时也将带来极高的成本。

Dojo采用特斯拉自研D1芯片



Dojo D1芯片微架构

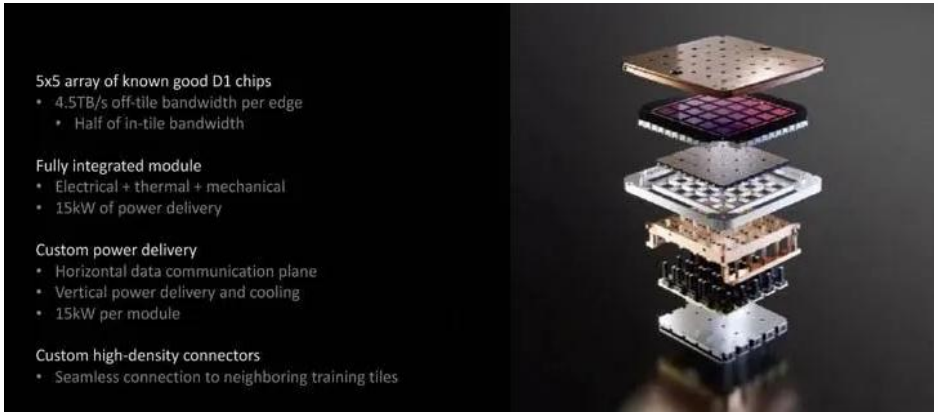


资料来源：Tesla AI Day，西南证券整理

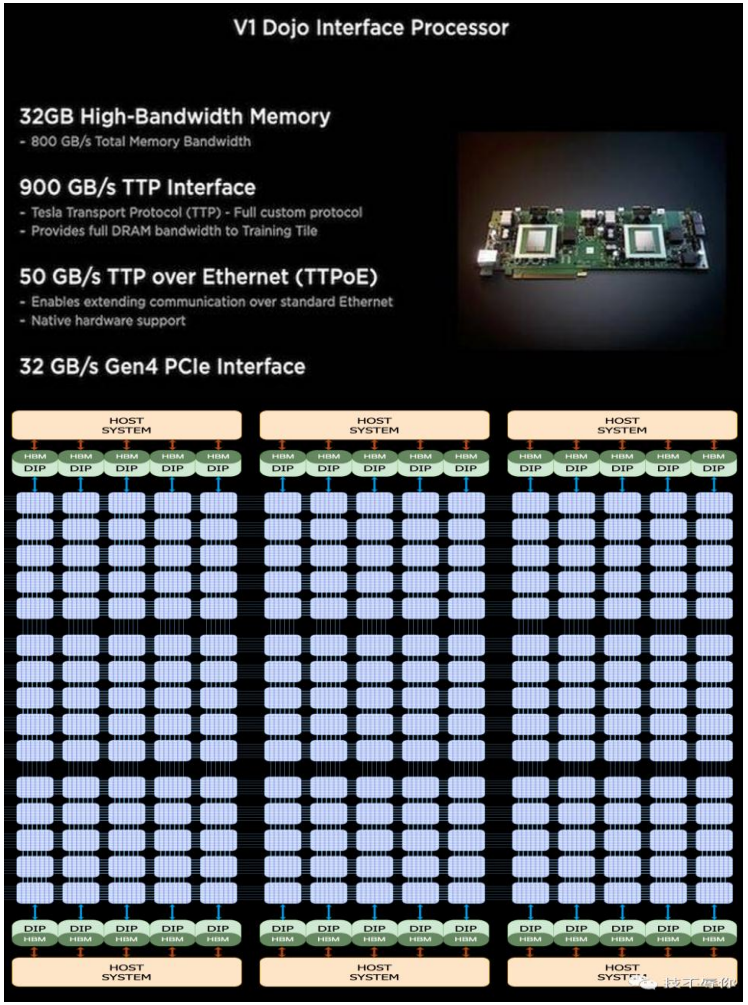
2.4 算力端：特斯拉Dojo——超高集成度的Training Tile

- **Training Tile**：基于D1芯片，特斯拉推出晶圆上系统级方案Training Tile，将计算、I/O、存储、液冷等模块高度集成，具备极低的延迟和极高的带宽。Training Tile应用台积电InFO_SoW封装技术，将5×5的D1芯片阵列排布组成，性能可以达到9PFlops BF16/CFP8，功耗为15KW。同时Tiles以2D Mesh结构互连，片上跨内核SRAM达到11GB，并在整个堆栈中使用特斯拉定制的传输协议，通过9TB/s结构连接。
- **Dojo接口处理器/Dojo Interface Processor**：Dojo接口处理器作为Tile与Host Server的通信桥梁，每个DIP提供900GB/s的TTP接口，同时配备32GB的HBM内存。每个Tile通过5张DIP卡与Host相连，则每个Tile的链路带宽达到4.5TB/s，可共享160GB的HBM。

25颗D1集成封装成为Dojo Training Tile



Training Tile通过DIP接口与主机连接

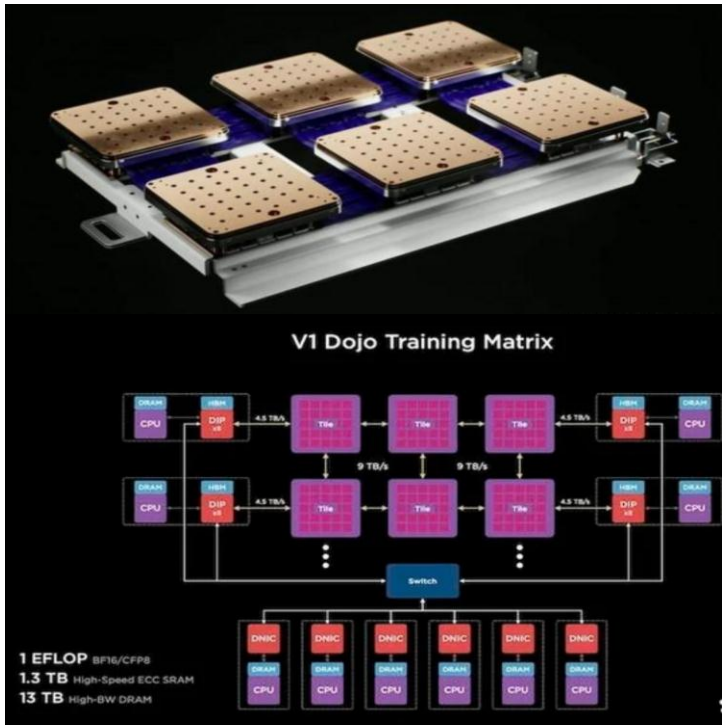


资料来源：Tesla AI Day，西南证券整理

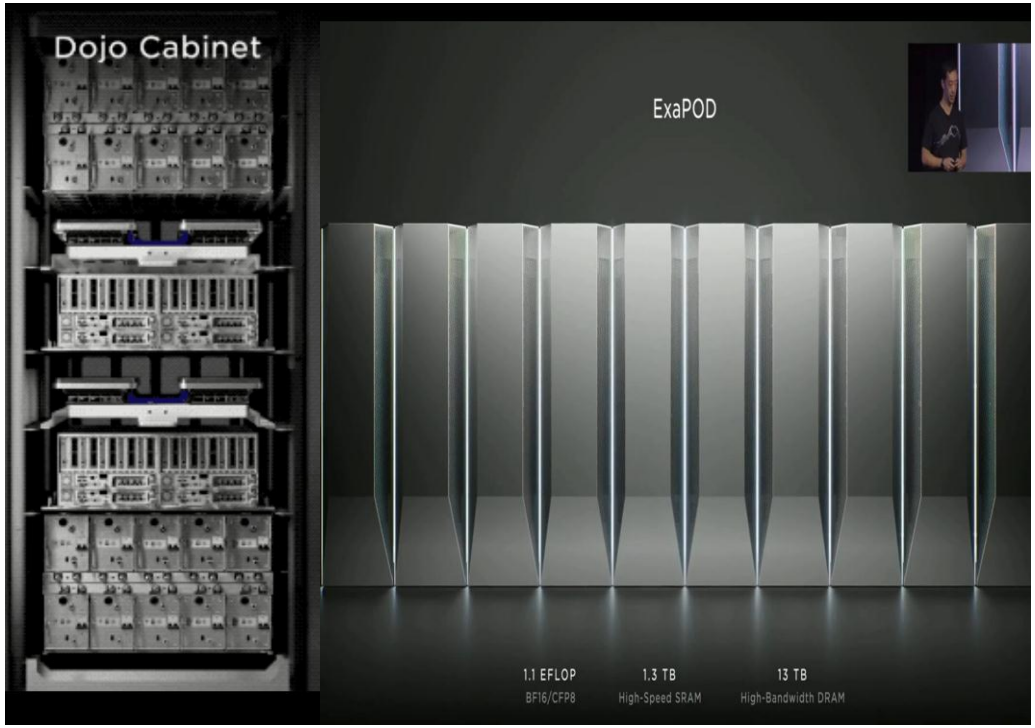
2.4 算力端：特斯拉Dojo——突破E级算力的ExaPOD

- 基于Training Tile，特斯拉推出ExaPOD大规模计算集群。
- 六块Training Tile组成一个Training Tray，单机柜可放置2个Tray；十个机柜组成一台ExaPOD。
- 1 ExaPOD=10 Cabinet=120 Tile=3000 D1 Chip=1062000 Node。
- 换算下来，一个ExaPOD可提供1.1EFlops算力，配备1.3TB SRAM，以及13TB DRAM。

六块Training Tile组成Training Tray



单机柜2个Tray，十个机柜组成ExaPOD

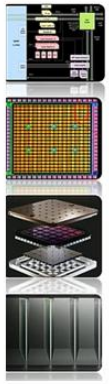


资料来源：Tesla AI Day，西南证券整理

2.4 算力端：特斯拉Dojo——架构设计的哲学

- **Dojo采用存算一体架构（存内计算or近存计算）。**
  - **面积精简**：将大量的计算内核集成到芯片中，最大限度提高AI计算的吞吐量，在保障算力的情况下使单个内核的面积尽可能小，更好地处理算力堆叠与延迟的矛盾。
  - **延迟精简**：为了实现区域计算效率最大化，内核以2GHz运行，只使用基本的分支预测器和小指令缓存，只保留必要的部件架构，其余面积留给向量计算和矩阵计算单元。
  - **功能精简**：通过削减对运行内部不是必须的处理器功能，进一步减少功耗和面积使用。Dojo核心不进行数据端缓存，不支持虚拟内存，也不支持精确异常。

Dojo硬件架构总结



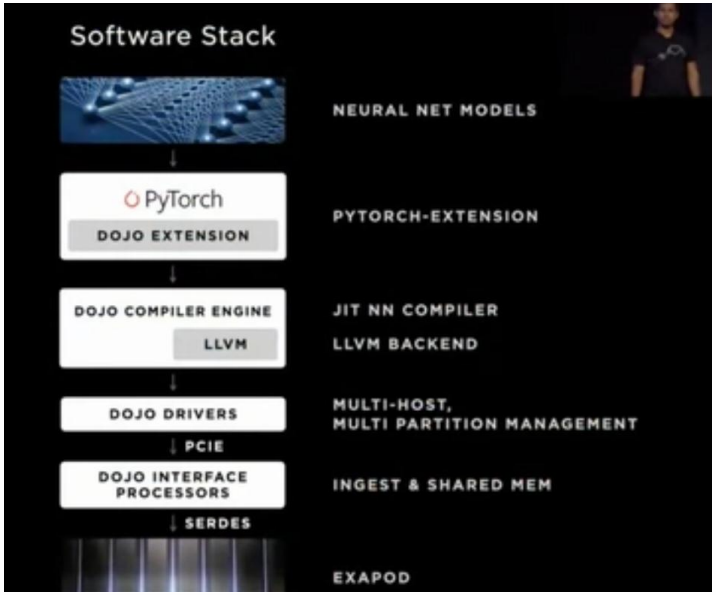
分层	名称	片上SRAM	算力（BF16、CFP8）	备注
内核	Dojo Node	1.25MB	1.024TFlops	单个计算核心，2GHz主频，4个8×8×4矩阵计算核心
芯片	Dojo D1	440MB	362TFlops	单芯片354核心，7nm，TDP为400W
模组	Dojo Tile	11GB	9950TFlops	5x5个D1芯片组成一个Tile
集群	Dojo ExaPOD	1320GB	1.1EFlops	12个训练模组组成一个机柜，10个机柜组成ExaPOD，共3000个D1芯片

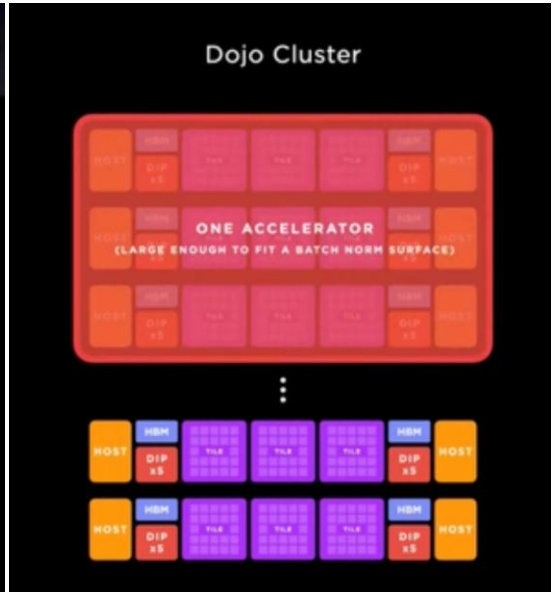
资料来源：Tesla AI Day，西南证券整理

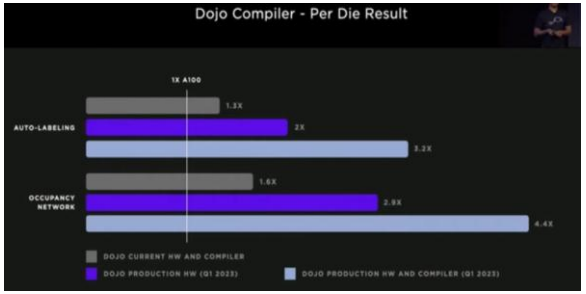
2.4 算力端：特斯拉Dojo——软件栈的优化

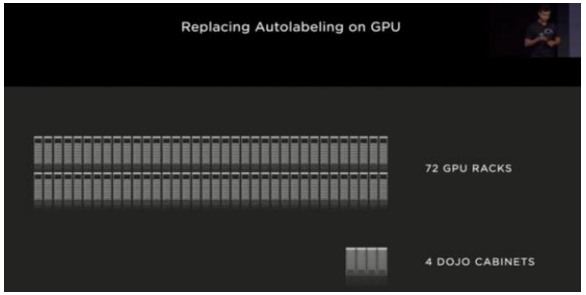
- 特斯拉Dojo的愿景是构建一个统一的加速器。
- Dojo硬件架构已经具备单个可扩展计算平面、全局寻址快速存储器和统一高带宽+低延迟的特性。
- 在此基础上，Dojo自建了编译器和指令集，ISA以Risc-V为基础进行改良与扩展，Dojo 编译器可以在尾数精度附近滑动，以涵盖更广泛的范围和精度。
- 通过软硬件层面的归一化，整个系统可以被抽象为一个整体进行算力调度，最大化硬件性能及利用率。根据特斯拉的测试，利用Dojo运行Occupancy Network神经网络模型时，相较英伟达A100能够实现性能数倍提升，过去需要6个GPU Box，现在只需要1个Dojo Tile。

软硬件层面归一化，带来的计算效率提升









资料来源：Tesla AI Day，西南证券整理

2.5 商业端：软件化进程推进，买断制叠加订阅制，整车价值量增加

- **从“量”的角度来看：**人工智能、神经网络以及大模型的应用正加快自动驾驶系统的迭代速度，技术的进步将带来用户驾驶体验的提升，从而推动用户付费转化。我们认为，特斯拉**FSD Beta v11.4版本在端到端大模型的赋能下将进一步优化系统性能，刺激软件需求量和付费率抬升。**
- **从“价”的角度来看：特斯拉FSD的收费模式采用买断制和订阅制。①买断制方面，**车主需要一次性支付套件价格，FSD从2016年的3000美元经过多轮涨价，自2022年9月5日起价格提升至15000美元。**②订阅制方面，**特斯拉在业内首创自动驾驶服务按月收费，FSD每月订阅价格在99美元至199美元之间，具体取决于车辆是否配备EAP系统；对于车主而言，订阅模式可以迅速降低FSD购买成本，并在使用期限上灵活选择；对于特斯拉而言，公司只需要开放软件接口即可增强盈利能力。**我们认为，无论是买断制还是订阅制，特斯拉在售卖整车的同时还具备软件价值，自动驾驶系统的迭代将增加整车价值量，电动汽车逐渐呈现软件化趋势。**

特斯拉驾驶系统价格复盘（美元）

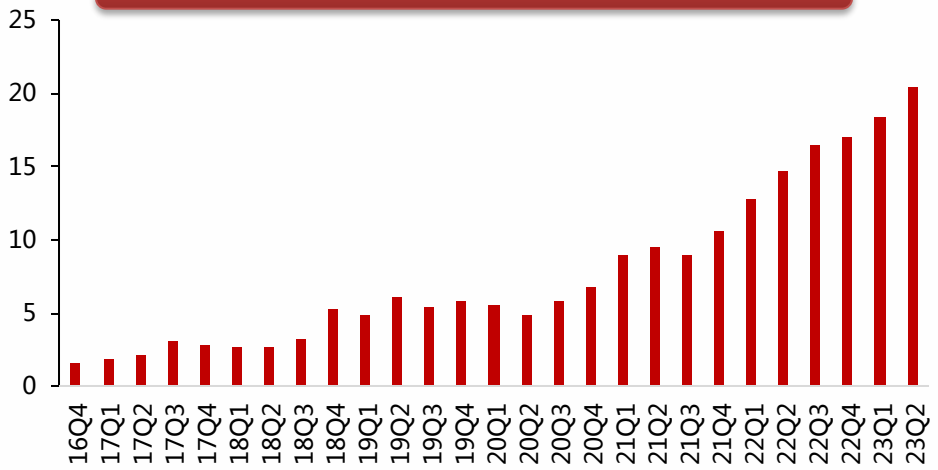
起始时间	终止时间	基础版自动辅助驾驶/AP	增强版自动辅助驾驶/EAP	全自动驾驶/FSD
2016年10月19	2019年2月27日	not available	5000	3000
2019年2月28	2019年4月10日	not available	3000	6000
2019年4月11日	2019年8月15日	included for free	not available	6000
2019年8月16日	2020年6月30日	included for free	not available	7000
2020年7月1日	2020年10月21日	included for free	not available	8000
2020年10月22日	2022年1月16日	included for free	not available	10000
2022年1月17日	2022年6月23日	included for free	not available	12000
2022年6月24日	2022年9月4日	included for free	6000	12000
2022年9月5日	至今	included for free	6000	15000

资料来源：特斯拉官网，西南证券整理

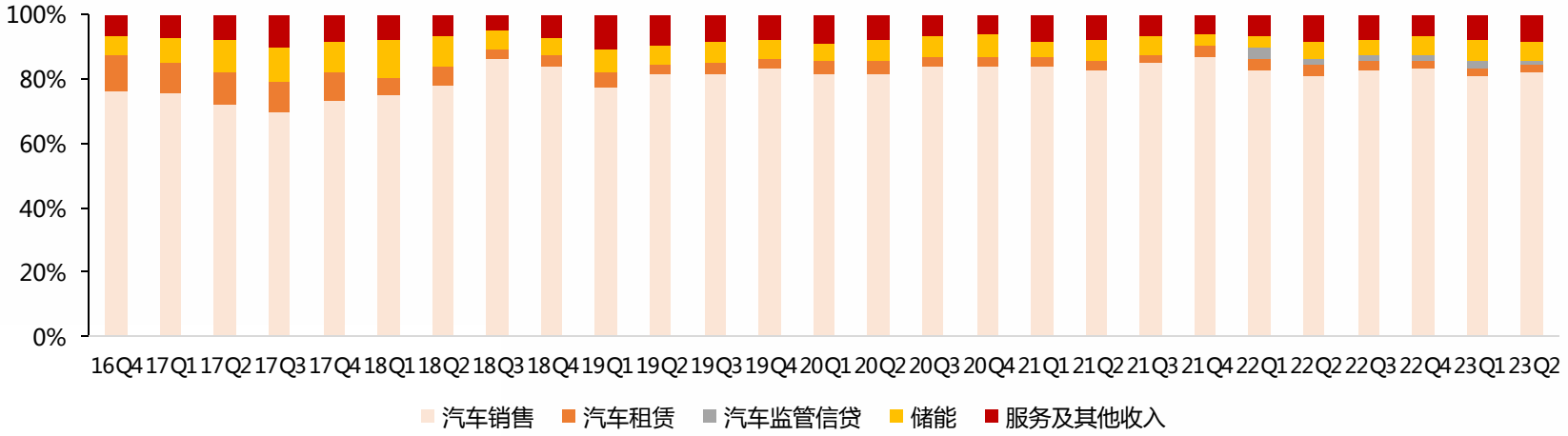
2.5 商业端：FSD套件业绩兑现，收入贡献日趋明显

- **FSD套件业绩兑现，收入贡献日趋明显。**随着FSD售价和搭载率提升，近年来FSD收入也随之上涨。根据公司2022年四季度财报，特斯拉FSD在22Q4带来约3.24亿美元收入。根据公司财报，包括FSD收入在内的服务及其他业务收入呈现持续增长态势，从16Q4的1.59亿美元增长至23Q2的20.5亿美元。我们认为，特斯拉FSD从2016年开始贡献收入，在业内率先产生业绩变现，未来智能汽车软件仍有巨大升值空间，特斯拉FSD汽车软件收入有望继续增长。

16Q4-23Q2特斯拉服务及其他收入（亿美元）



16Q4-23Q2特斯拉收入结构



资料来源：Wind，西南证券整理

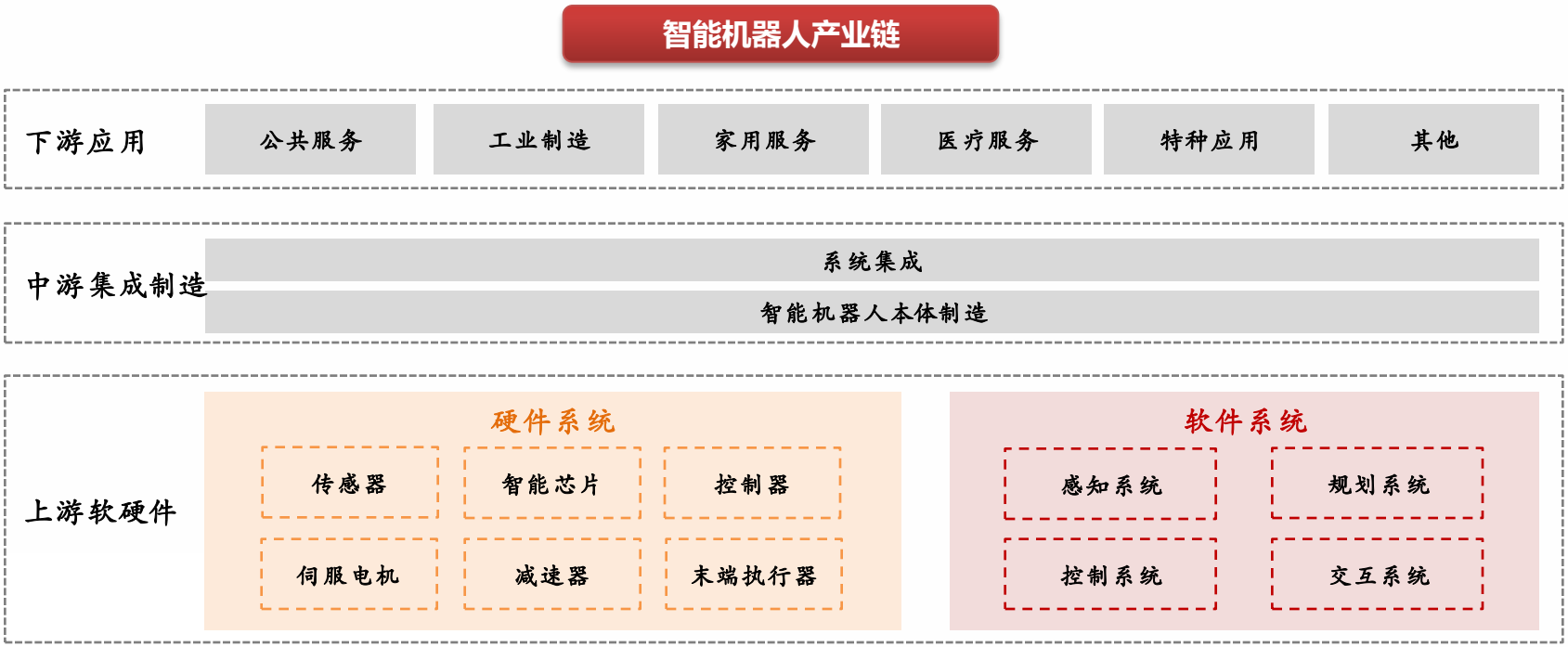
目 录

3 AI赋能人形机器人，引领具身智能浪潮

- **机器人作为具身智能的更优形态，人工智能将对机器人进一步赋能。**

- 具身智能是将人工智能与机器结合，将多模态的大语言模型作为人类与机器沟通的桥梁，帮助机器处理具身推理任务，强调智能与身体和环境的互动关系，将智能与实际物理世界结合起来，通过身体感知、运动和与环境互动来实现智能行为。

- 人形机器人的具身智能包括具身感知和具身执行。其中，具身感知是指通过机器人身上的各种传感器获取周围环境的信息。具身执行是指将机器人的感知和决策转化为具体行动。近年来，人形机器人作为具身智能的代表产品，结构设计日益符合人类特点，**AI技术的进步进一步提升了人形机器人的感知、规划、控制和人机交互能力。**



资料来源：甲子光年，西南证券整理

3 AI赋能人形机器人，引领具身智能浪潮

- **AI技术的进步进一步提升了人形机器人的语音能力、视觉能力和运动能力，但同时也存在众多难点。**
- **硬件核心难点：人形机器人在手部和腿部的硬件集成上难度较大。**1）手部：手部集成需要大量的电机和驱动器；2）腿部：人形机器人的腿部驱动器需要很高的损失峰值功率和驱动能力。
- **软件核心难点：**1）手部：手部涉及20多个自由度，**精细化感知**难度大；2）腿部：人形机器人在腿部行走上尚未出现真正意义上的**类人行走算法**，行走算法的技术难度大，且当前的行走算法有很大的不稳定性。3）全身的结合：全身控制需要结合躯干、双臂和腿部等，涉及到难度较大的**复合算法**。4）规控算法：**混合智能操控和避让**等算法难度大，需要**更高维度的规划。**

主流人形机器人性能对比

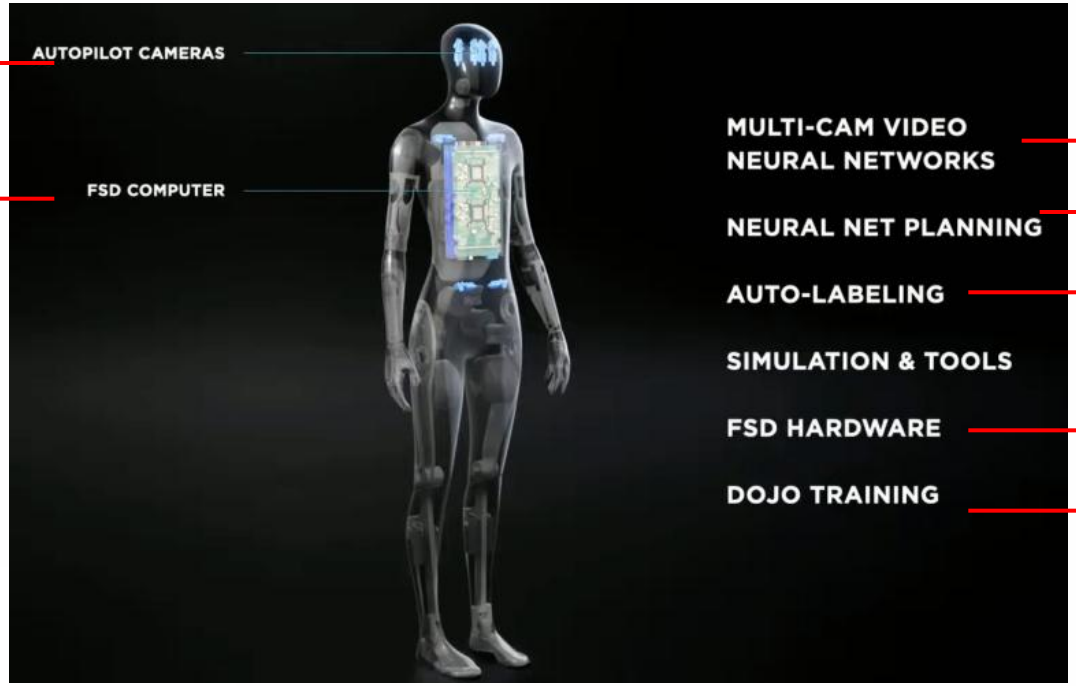
	特斯拉Optimus	波士顿动力Atlas	优必选Walker X	小米CyberOne
身高	1.8米	1.75米	1.3米	1.78米
重量	57kg	75kg	63kg	52kg
语音能力	Tesla SoC：语音交流	/	四维灯语体系、语音交互	85种环境语义识别；6类45种人类语义情绪识别
视觉能力	Tesla SoC：视觉信息处理；Autopilot的摄像头作为视觉传感器，共8个摄像头	两个视觉系统：1个激光测距仪+1个立体照相机	U-SLAM视觉导航技术，实现自主规划路径；定位精度10cm；精定位精度1cm；导航精度20cm	2D弯曲；OLED屏幕；Mi Sense自研空间视觉；AI交互相机
运动能力	速度8km/h；身体28个自由度，手部11个自由度；驱动：电机+减速器；承载最多约20kg的物品；可行走、上下楼梯、下蹲、拿取物品等动作	速度1.5m/s；四肢28个自由度；驱动：电机与液压两种传动；可垂直起跳、跨越障碍、后空翻等	速度3km/s；四肢41个自由度；驱动：电机+减速器；可在20°斜坡上行走，15cm台阶上上下楼梯；动态足腿控制，自平衡抗干扰	速度3.6km/s；21个关节自由度；驱动：电机+减速器
应用场景	特种机器人：为人类执行一些无聊或危险的任务，如搬运重物、采购杂货等	特种机器人：执行巡逻、勘测、运输任务等	服务型机器人	服务型机器人

资料来源：中移智库，西南证券整理

3 特斯拉横向迁移FSD底座，机器人与自动驾驶软硬件部分适用

- **硬件层面：特斯拉自动驾驶和机器人在硬件上具备一定的通用性。**感知层主要包括摄像头、毫米波雷达等传感器；规划层主要基于AI芯片和FSD系统；控制层包括执行器等。特斯拉机器人在硬件端与自动驾驶具有一定相似性。
- **软件层面：特斯拉打通FSD在自动驾驶和机器人中的底层模块，在一定程度上实现算法的复用。**自动驾驶FSD系统可以根据感知到的环境信息进行路径规划和车辆控制，该方法同样适用于机器人，帮助机器人实现视觉感知、从而在复杂环境中选择最佳路径、最后执行适当的决策。实际上，自动驾驶本质也属于机器人，特斯拉目前在感知和识别等模块上具有一定的通用人工智能能力，而通用人工智能算法将是特斯拉未来长期价值所在。

特斯拉通用机器人的人工智能



基于AutoPilot的摄像头

复用FSD算力设施

基于神经网络算法

神经网络规划

复用自动标注系统

复用FSD硬件设施

复用Dojo训练平台

资料来源：Tesla AI Day，西南证券整理

3.1 硬件端-视觉传感器：坚持纯视觉路线，基于多目打造立体感知

- **机器人的智能感知离不开视觉传感器，**视觉传感器主要用于检测机器人周围的环境，并转化为机器人可以理解的数据和信息。
- **机器人视觉主要分为2D视觉和3D视觉：1）2D视觉**主要基于摄像头和距离传感器进行感知、并通过算法还原深度数据，硬件成本低，但算法难度大；**2）3D视觉**可分为激光雷达和深度相机，**深度相机**又分为双目RGB相机、结构光相机和TOF相机，主要用于检测空间的景深距离。双目RGB相机非常依赖纯图像特征的提取和匹配，是纯视觉方法，因此对场景的光照和纹理要求较高、对算法要求高、计算量较大，应用场景主要为双目视觉搬运机器人/机械臂、双目扫地机器人等；而结构光相机可在光照不足、缺乏纹理的场景使用，例如3D人脸识别、手势识别、安全验证、金融支付等场景；TOF飞行时间法通过发射持续不断的“面光源”，快速计算与物体的距离，得到被测物体的3D图像，可用于机器人导航、规划路径、实现避障等场景。

机器人视觉传感器技术指标对比

指标	激光雷达	摄像头	毫米波雷达	超声波雷达	红外
远距离探测能力	强	强	强	弱	一般
夜间工作能力	强	弱	强	强	强
受气候影响	大	大	小	小	大
烟雾环境工作	弱	弱	强	一般	弱
雨雪环境工作	一般	一般	强	强	弱
温度稳定性	强	强	强	弱	一般
车速测量能力	弱	弱	强	一般	弱
行人测量能力	一般	强	弱	弱	弱
测量精度	高	中	中	低	高
分辨率	高	中	中	低	低
成本	高	中	中	低	较低

机器人视觉摄像头性能和成本比较

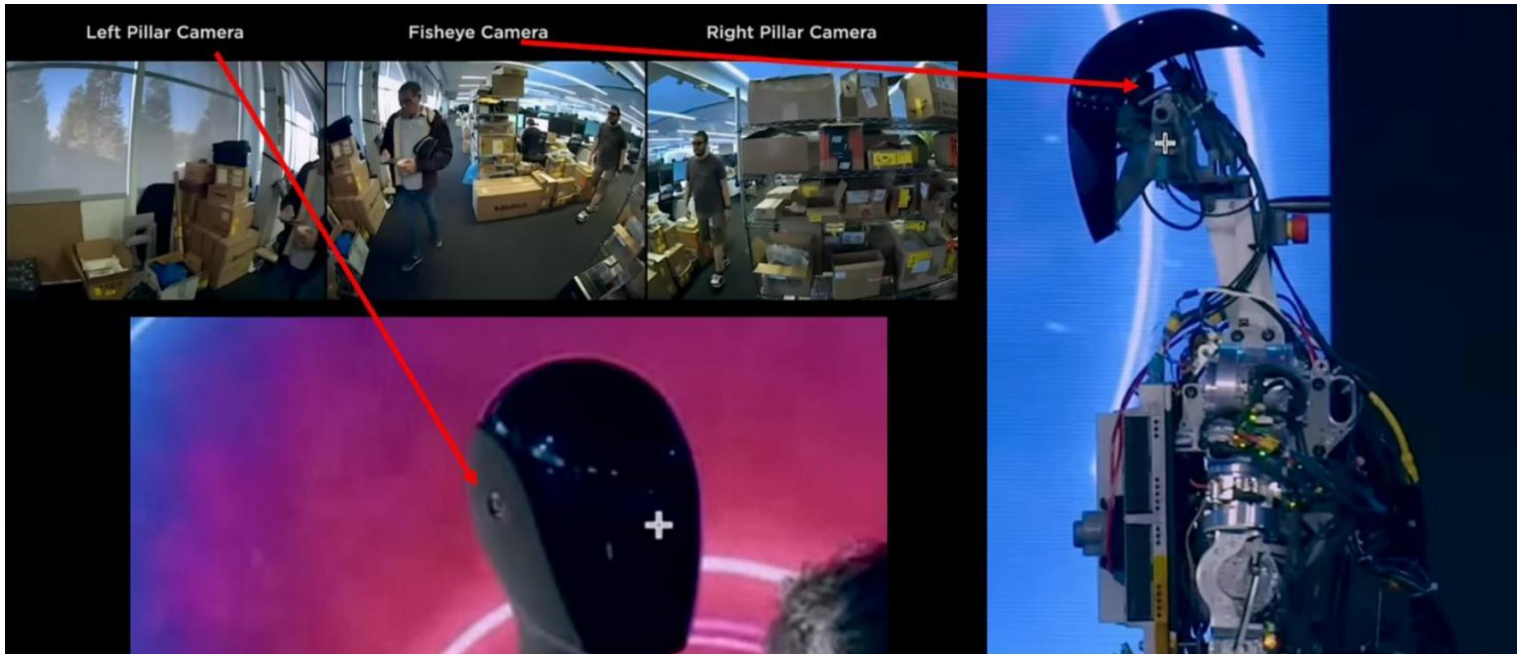
指标	普通单目	单目结构光	单目TOF	双目结构光	普通双目
检测距离	远	近	近	近	远
分辨率	高	中	低	中	高
精度	低	高	高	高	中
视角	广	窄	窄	窄	广
帧率	高	低	中	低	中
功耗	低	高	高	高	中
成本	低	高	高	高	中
适应环境	室内室外	室内	室内室外	室内室外	室内室外

资料来源：焉知汽车，西南证券整理

3.1 硬件端-视觉传感器：坚持纯视觉路线，基于多目打造立体感知

- **视觉感知方面，特斯拉坚持纯视觉路线，依靠摄像头采集信息。**自动驾驶和机器人均通过传感器获取周围环境信息，常用传感器包括摄像头、雷达（毫米波/激光/超声波等）、红外传感器、GPS、IMU等。在主流机器人中，特斯拉Optimus沿用自动驾驶感知方案，采用纯视觉路线；波士顿动力Altas机器人则采用多传感器路线，包括激光测距仪和立体照相机两个视觉系统。
- **基于双目摄像头视差原理，打造立体视觉感知。**特斯拉Optimus机器人头部配置3颗Autopilot摄像头，包括左肩摄像头、右肩摄像头和中央鱼眼摄像头，可覆盖大于180度的体前场景。特斯拉采用双目摄像头，其原理与人眼相似，基于视差使视觉感知更加立体，且双目系统成本与激光雷达方案相比成本更低。

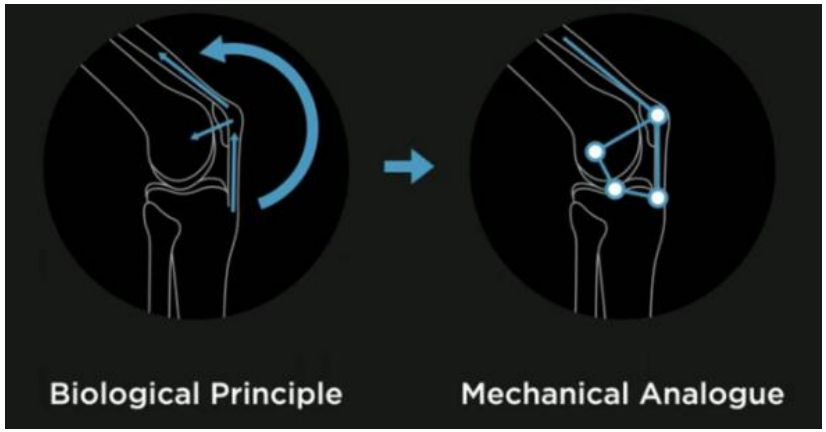
特斯拉Optimus头部摄像头配置



资料来源：Tesla AI Day，西南证券整理

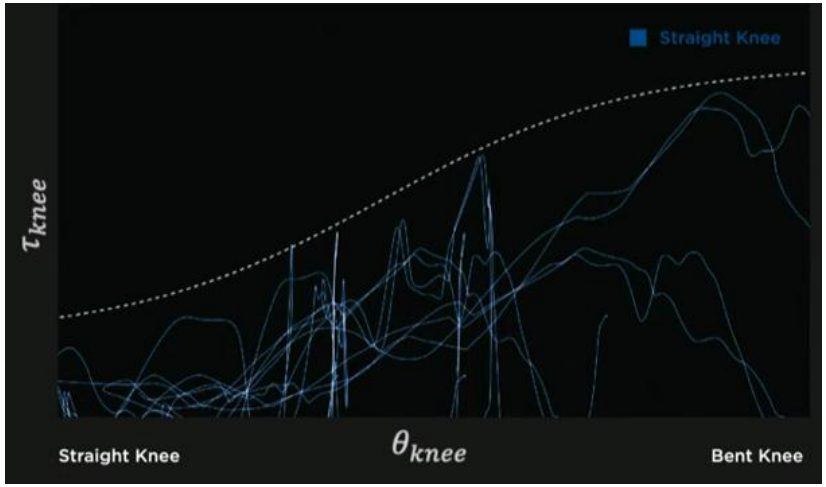
3.1 硬件端-四连杆膝盖关节：模拟人体设计，优化腿部力学模型

基于仿生学的四连杆膝盖关节

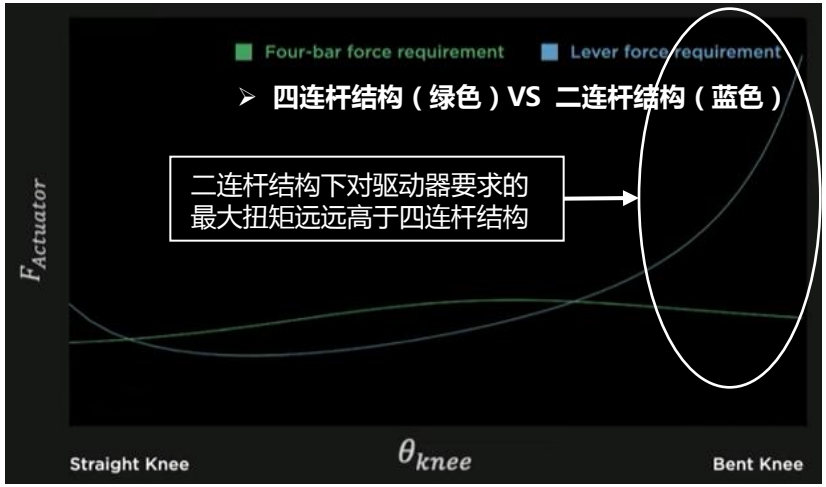


- **腿部膝盖弯曲角越大，膝部扭矩负载更高。**随着腿部的弯曲角度变大，例如越接近蹲姿，执行同一任务所需的扭矩会越来越大，例如半蹲走路比站着走路更费劲。
- **特斯拉采用四连杆结构，让同一负载在直腿状态和弯腿角度下的所需扭矩更为平缓和一致。**在简单的二连杆设计结构下，机器人的大小腿仅用一个转轴连接，导致机器人在弯腿状态下所需的执行扭矩会显著增加（蓝线）；在四连杆膝盖结构下，所需扭矩基本保持平稳，将实现小马拉大车的效果。

膝部负载扭矩和弯曲角关系



执行扭矩图谱



资料来源：Tesla AI Day，西南证券整理

3.1 硬件端-驱动器选型：基于成本-轻量化考虑，实现一机多用

- **具备一机多用的通用性：**特斯拉Optimus躯干中共有28个关节，28各关节中总共采用6种驱动器，实现一机多用，具备通用性，避免由于系统内驱动器型号过多而导致的生产制造效率低下、成本较高等问题。
- **具备更省力的扭重比：**6个驱动器中包括3个旋转驱动器和3个直线驱动器，每个驱动器均有较好的扭重比（即发动机作用于每1吨重量的扭矩，单位为NM），在一定体积下具备更大力气。

机器人躯干中的28个关节共采用6类驱动器

资料来源：Tesla AI Day，西南证券整理

3.1 硬件端-驱动器配置：机械结构输出大力矩，传感器助力精准电控

- 特斯拉Optimus在旋转驱动器和直线驱动器上**十分注重扭矩的输出**，例如直线驱动器内的倒置滚珠丝杠。特斯拉在传感器上**致力于实现更精准的电控**，在驱动器内部均布置相应的位置传感器和力矩传感器。

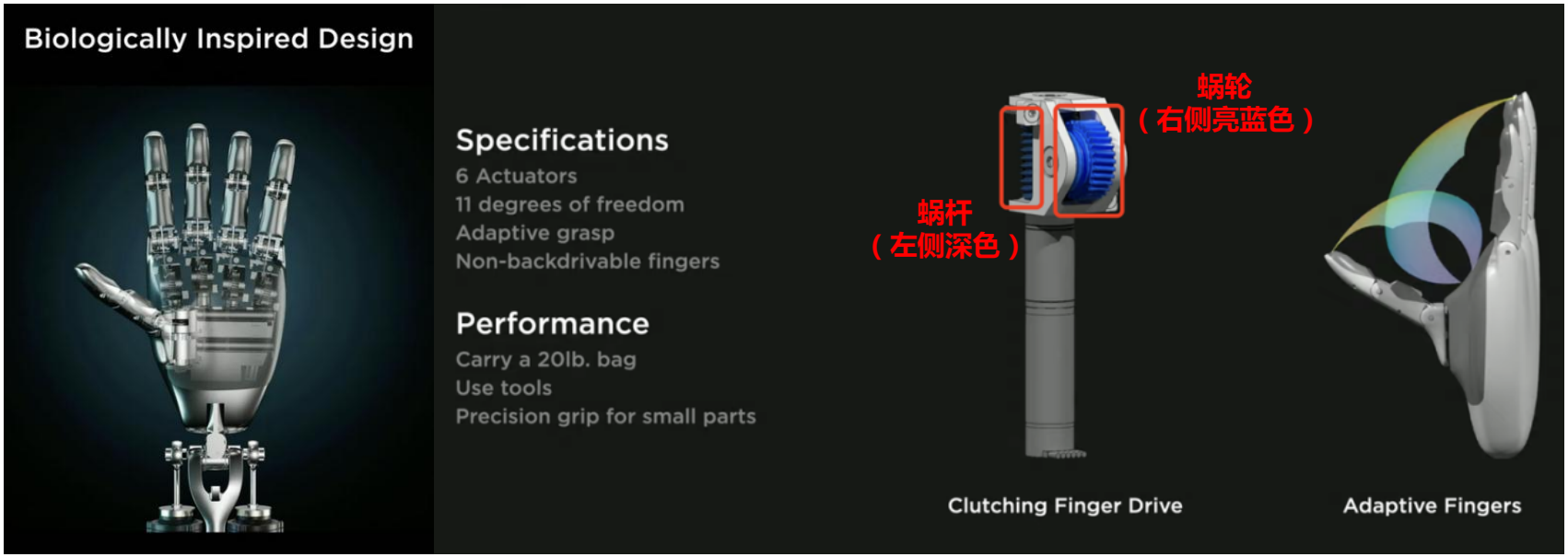
三类旋转驱动器和三类直线驱动器的具体硬件配置

资料来源：Tesla AI Day，西南证券整理

3.1 硬件端-机器手：采用电机驱动方案，追求灵巧且高效

- 特斯拉Optimus机器手**基于仿生学关节设计，模拟人类关节与肌腱形态**，每只手共搭载6个驱动器，拥有11个自由度（6个主动自由度+5个被动自由度），并搭载传感器，**具备感知功能和自适应抓取能力**，机器手通过对不同物体的抓握进行模型训练逐步形成手部的触觉感知，目前，Optimus可提起约重9公斤的物品、使用部分工具、并能够精确抓取一些小部件。
- 机器手的手指关节采用具备自锁结构的**蜗轮蜗杆设计，实现轻量且高效。**指关节由蜗杆驱动蜗轮，但蜗轮无法驱动蜗杆，由此避免因关节负重而导致驱动器反转，同时在提取重物时关节会因自锁效应固定从而保持手部姿势、避免指关节驱动器额外工作。

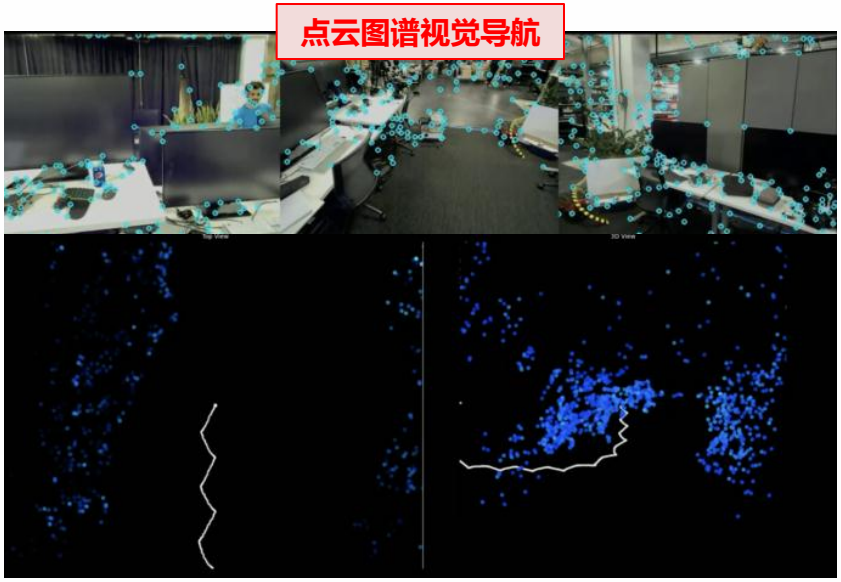
机器手采用人体仿生设计 & 手指关节采用的蜗轮蜗杆设计



资料来源：Tesla AI Day，西南证券整理

3.2 算法端-感知：复用底层算法，改进占用网络，实现视觉导航

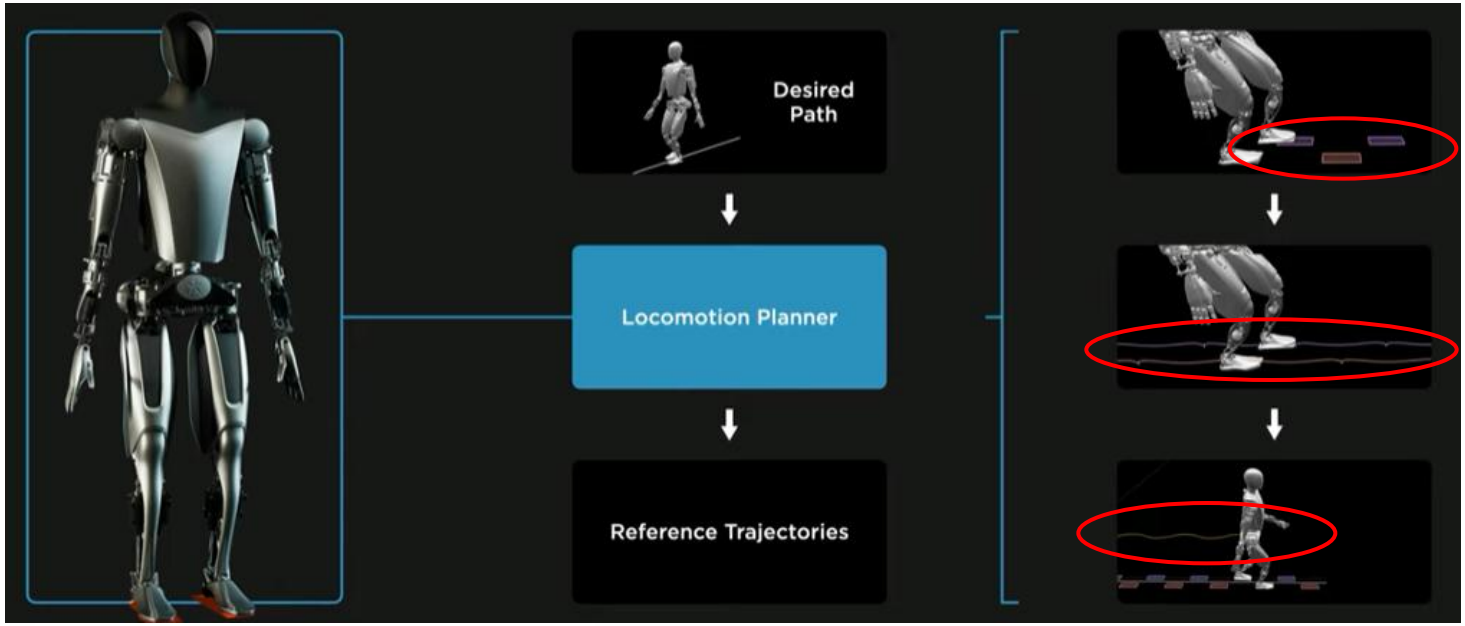
- **FSD打通自动驾驶和机器人的底层模块，复用视觉神经网络。**特斯拉机器人所运用的视觉神经网络直接由自动驾驶系统移植而来。其中，占用网络仍是重中之重。在3D实物探测中，特斯拉进一步改进占用网络，使用NeRF完成图形的3D渲染，在视觉感知上强调提供更加精确的景深和体积感知。
- **构建空间点云图谱，实现机器人视觉导航。**在路径导航方面，特斯拉将机器人通过视觉检测到的物体搭建成一个空间点云图谱，通过训练让机器人识别室内环境下的常见物体和关键特征，然后再在图谱中避开环境中的实体从而规划出可行进路径。



资料来源：Tesla AI Day，西南证券整理

3.2 算法端-规划：借用自动驾驶模拟器，融合多学科，优化运动轨迹

- **自动驾驶模拟器可执行机器人运动代码，但机器人移动较汽车移动更加复杂。**在模拟方面，特斯拉将机器人的运动代码集成到自动驾驶模拟器中，通过运行自动驾驶模拟器的运动控制代码，帮助机器人实现行走。2022年4月，特斯拉机器人迈出第一步，移动速度缓慢；但随着团队解锁更多关节、以及技术的不断进展，例如手臂平衡等，机器人的行走日益进化。事实上，从汽车移动到机器人移动的过程中，运动规划变得更加深入和复杂。
- **模型基于多种学科，优化路径和轨迹规划。**人类在行走过程中具备身体的自我意识、采用节能步态、能够做到平衡和四肢协调，因此，机器人的运动规划需要结合运动学、动力学和接触特性等多种学科，模型更加复杂。当前，机器人的行走规划主要分为三个部分：1）基于期望路径，进行足迹规划；2）基于规划的足迹，将足迹连接成轨迹，机器人通过脚趾和脚跟的步幅在轨迹上实现行走，提供更大的步幅和更少的膝盖弯曲，从而提高系统效率；3）找到质心轨迹，保证机器人系统在平衡状态下进行动态的、可行的运动。



➢ **第一步：**
基于期望路径，进行足迹规划

➢ **第二步：**
将足迹连接成轨迹，机器人基于轨迹、并运用脚趾和脚跟像人类一样跨步走

➢ **第三步：**
找到质心轨迹，保证机器人在平衡状态下进行动态的、可行的运动

资料来源：Tesla AI Day，西南证券整理

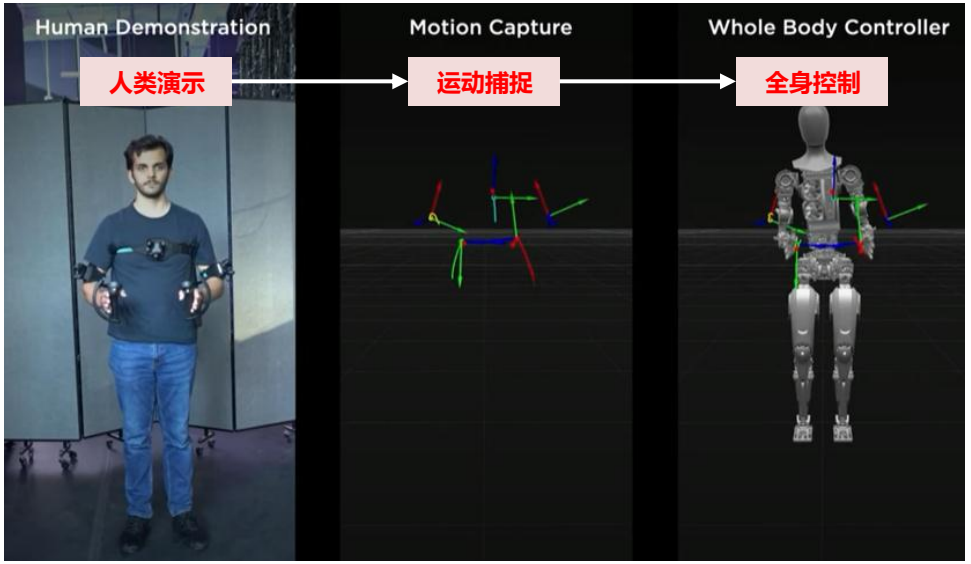
3.2 算法端-控制：学习人类动作，添加轨迹优化程序，适应现实世界

- 特斯拉机器人团队为实现机器人在现实世界中更加自然地操纵事物，主要基于以下两个步骤：

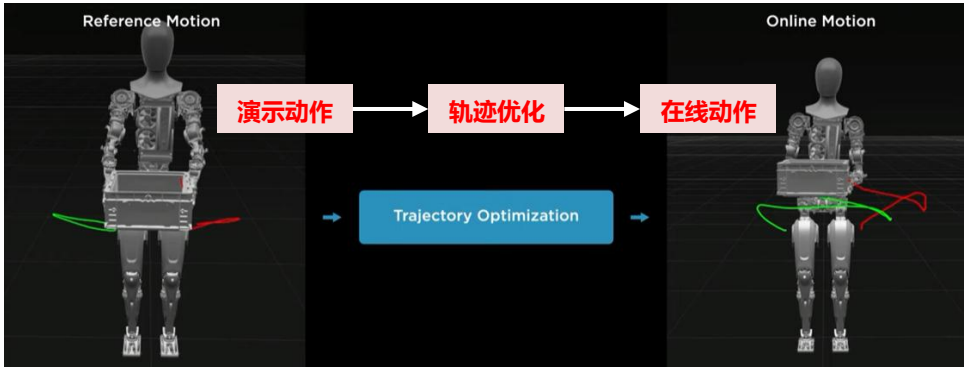
- **1）生成自然运动参考库：**通过收集**人类演示**的自然动作，形成可供机器人参考和学习的运动库。**机器人动作学习过程具体如下：**→①人类演示拾取物体的动作和过程→②系统对人类的动作进行捕捉→③将动作可视化地转化成手部、肘部和躯干位置的关键帧→④使用反向运动学将其映射到机器人上。

- **2）在线运动自适应：**在第一个步骤中，单一动作的演示并不足以**适应现实世界中的变化**，因此特斯拉推出在线运动自适应技术。例如，以机器人拾取特定位置的长方体为例，**特斯拉通过添加轨迹优化程序，帮助解决手应该在哪里、机器人应该如何平衡，何时需要将运动适应现实世界等问题。最终生成可以适应现实世界变化的运动参考轨迹。**

➢ **第一步：自然运动参考库（Natural Motion References）**



➢ **第二步：在线运动自适应（Online Motion Adaptation）**



资料来源：Tesla AI Day，西南证券整理

风险提示

- 行业竞争加剧风险
- 相关技术发展不及预期风险
- 商业变现不及预期风险

西南证券
SOUTHWEST SECURITIES

分析师：王湘杰
执业证号： S1250521120002
电话：0755-26671517
邮箱：wxj@swsc.com.cn

西南证券 SOUTHWEST SECURITIES

西南证券研究发展中心

西南证券投资评级说明

报告中投资建议所涉及的评级分为公司评级和行业评级（另有说明的除外）。评级标准为报告发布日后6个月内的相对市场表现，即：以报告发布日后6个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A股市场以沪深300指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克综合指数或标普500指数为基准。

公司评级	买入：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在20%以上
	持有：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于10%与20%之间
	中性：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-10%与10%之间
	回避：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-20%与-10%之间
	卖出：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在-20%以下
行业评级	强于大市：未来6个月内，行业整体回报高于同期相关证券市场代表性指数5%以上
	跟随大市：未来6个月内，行业整体回报介于同期相关证券市场代表性指数-5%与5%之间
	弱于大市：未来6个月内，行业整体回报低于同期相关证券市场代表性指数-5%以下

分析师承诺

报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接获取任何形式的补偿。

重要声明

西南证券 SOUTHWEST SECURITIES

西南证券研究发展中心

西南证券研究发展中心

上海

地址：上海市浦东新区陆家嘴21世纪大厦10楼

邮编：200120

北京

地址：北京市西城区金融大街35号国际企业大厦A座8楼

邮编：100033

深圳

地址：深圳市福田区益田路6001号太平金融大厦22楼

邮编：518038

重庆

地址：重庆市江北区金沙门路32号西南证券总部大楼21楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	手机	邮箱	姓名	职务	手机	邮箱
上海	蒋诗烽	总经理助理/销售总监	18621310081	jsf@swsc.com.cn	张玉梅	销售经理	18957157330	zymyf@swsc.com.cn
	崔露文	销售经理	15642960315	clw@swsc.com.cn	陈阳阳	销售经理	17863111858	cyyyf@swsc.com.cn
	谭世泽	销售经理	13122900886	tsz@swsc.com.cn	李煜	销售经理	18801732511	yfliyu@swsc.com.cn
	薛世宇	销售经理	18502146429	xsy@swsc.com.cn	卞黎旸	销售经理	13262983309	bly@swsc.com.cn
	刘中一	销售经理	19821158911	lzhongy@swsc.com.cn	龙思宇	销售经理	18062608256	lsyu@swsc.com.cn
	岑宇婷	销售经理	18616243268	cyryf@swsc.com.cn	田婧雯	销售经理	18817337408	tjw@swsc.com.cn
	汪艺	销售经理	13127920536	wyyf@swsc.com.cn	阚钰	销售经理	17275202601	kyu@swsc.com.cn
北京	李杨	销售总监	18601139362	yfly@swsc.com.cn	姚航	销售经理	15652026677	yhang@swsc.com.cn
	张岚	销售副总监	18601241803	zhanglan@swsc.com.cn	胡青璇	销售经理	18800123955	hqx@swsc.com.cn
	杨薇	高级销售经理	15652285702	yangwei@swsc.com.cn	王宇飞	销售经理	18500981866	wangyuf@swsc.com
	王一菲	销售经理	18040060359	wyf@swsc.com.cn	路漫天	销售经理	18610741553	lmtyf@swsc.com.cn
	徐铭婉	销售经理	15204539291	xumw@swsc.com.cn	马冰竹	销售经理	13126590325	mbz@swsc.com.cn
广深	郑龑	广深销售负责人	18825189744	zhengyan@swsc.com.cn	张文锋	销售经理	13642639789	zwf@swsc.com.cn
	杨新意	销售经理	17628609919	yxy@swsc.com.cn	陈紫琳	销售经理	13266723634	chzlyf@swsc.com.cn
	龚之涵	销售经理	15808001926	gongzh@swsc.com.cn	陈韵然	销售经理	18208801355	cyryf@swsc.com.cn
	丁凡	销售经理	15559989681	dingfyf@swsc.com.cn				