

- **何谓“大模型”？大数据喂养神经网络算法，设定规则机制使其自成长。** 1) 多模态数据是基础，包括文本/图像/音频/视频等在内的多类型数据喂养，驱动算法更好完成理解/生成等任务。2) **Transformer是核心**，Self-Attention机制强化算法抽象特征提取能力，并支持并行计算，高能且高效，衍生ViT/DiT支持多模态数据理解/生成。3) **ChatGPT及Sora为代表应用**，在大参数加持下，开发多模态自然语言处理以及文生视频等功能。4) **世界模型为未来方向**，算法能力由数据驱动演变为认知驱动，模型具备反事实推理和理解物理客观规律的能力，提升通用泛化特性。
- **大模型重塑车端算法架构，加速云端算法迭代；世界模型或为完全自动驾驶最优解。** 智驾一阶段（L2~L3）**脱胎换骨**：由场景驱动转向数据驱动，大模型带来底层架构质变；智驾二阶段（L3~L4）**厚积薄发**：由数据驱动转向认知驱动，数据和算力逐步累计驱动能力提升，量变引起质变。1) **车端**：上层感知/规控应用层算法随功能需求提升持续进化，“场景理解处理能力泛化”是核心诉求。L2~L3，感知端为升级核心，Transformer加持BEV+占用网络算法落地感知端到端，解决长尾场景识别难题；L3~L4以规控算法升级为核心，精准识别并快速处理，Learning-base逐步取代Rule-base，端到端拉高场景处理能力的天花板。2) **云端**：数据闭环为前提，加速大数据有效利用，采集/标注/仿真/训练/部署一体化。Transformer赋能自动标注，数据驱动场景仿真泛化，降低对有限的实际路测数据的依赖。3) **世界模型【通用具身智能】或为自动驾驶最优解。**车端场景生成泛化，将自动驾驶问题转化为预测视频的下一帧，类人模式处理，实现泛化至数据场景以外的能力；并可快速生成标准化仿真数据，结合大算力加速云端训练。
- **算法实现高壁垒+数据闭环硬要求，未来智驾算法产业格局趋于集中化。**当前下游L3+高阶算法方案以OEM自研为主，华为以“算法赋能，数据回传”的形式深度绑定OEM；L3以下算法呈现OEM+独立算法商+硬件商三足鼎立格局。我们认为，考虑高阶智驾功能的提升对于算法能力/组织架构/超算中心/完整数据链等的要求，未来“**掌握硬件的基础上去发展软件**”或为主流，即**掌握壁垒最高的硬件——芯片；提供性价比最高的硬件——传感器；掌握粘性最强的硬件——整车。**

■ 投资建议：汽车AI智能化转型大势所趋，算法为主干，看好头部算法玩家持续领先铸就高壁垒。

- **全行业加速智能化转型，产业趋势明确。** 下游OEM玩家+中游Tier供应商均加大对汽车智能化投入，大势所趋；智驾核心环节【软件+硬件+数据】均围绕下游OEM展开，数据催化算法提效进而驱动硬件迭代。以特斯拉为代表，应用算法向全栈端到端-世界模型持续迭代，功能落地兑现。
 - OEM整车厂商&核心芯片硬件厂商&智驾传感器厂商&独立算法商加速布局端到端算法开发，场景驱动-数据驱动-认知驱动持续进化；智驾算法产业发展进入深水区，高投入赋能【大算力+大数据】，**方能走通L3有条件自动驾驶至L4完全自动驾驶之路。**
 - **看好智驾头部车企以及智能化增量零部件：** 1) 华为系玩家【长安汽车+赛力斯+江淮汽车】，关注【北汽蓝谷】； 2) 头部新势力【小鹏汽车+理想汽车】； 3) 加速转型【吉利汽车+上汽集团+长城汽车+广汽集团】； 4) **智能化核心增量零部件：** 域控制器（德赛西威+经纬恒润+华阳集团+均胜电子等）+线控底盘（伯特利+耐世特+拓普集团等）。
- **风险提示：** 智能驾驶相关技术迭代/产业政策出台低于预期；华为/小鹏等车企新车销量低于预期。



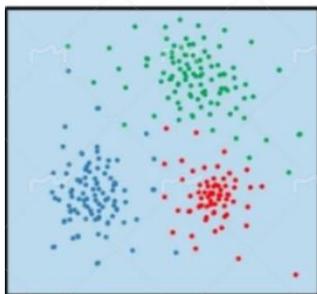
- 一、何谓“大模型”？
- 二、车端：大模型重塑智驾算法架构
- 三、云端：大模型加速智驾算法迭代
- 四、当前产业玩家的智驾算法能力如何？
- 五、投资建议与风险提示

一、何谓“大模型”？

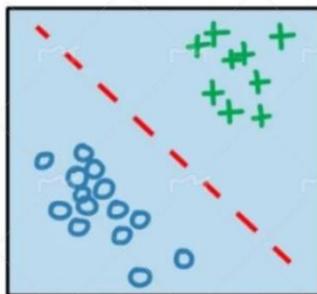
图：机器学习

machine learning

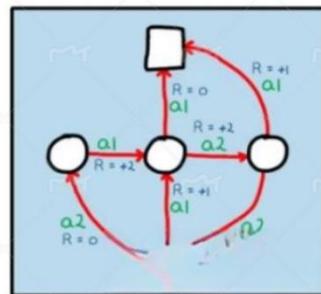
unsupervised learning



supervised learning



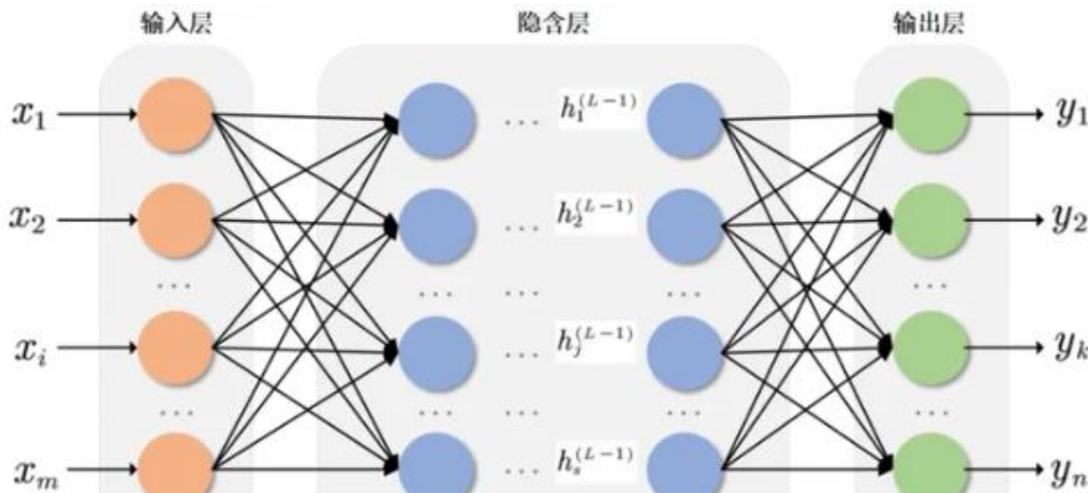
reinforcement learning



机器学习 > 神经网络 > 深度学习 ≈ 深度神经网络。机器学习用于解决由人工基于 if-else 等规则开发算法而导致成本过高的问题，想要通过帮助机器“发现”它们“自己解决问题的算法”来解决；机器学习可以分为有监督学习、无监督学习和强化学习等三类。

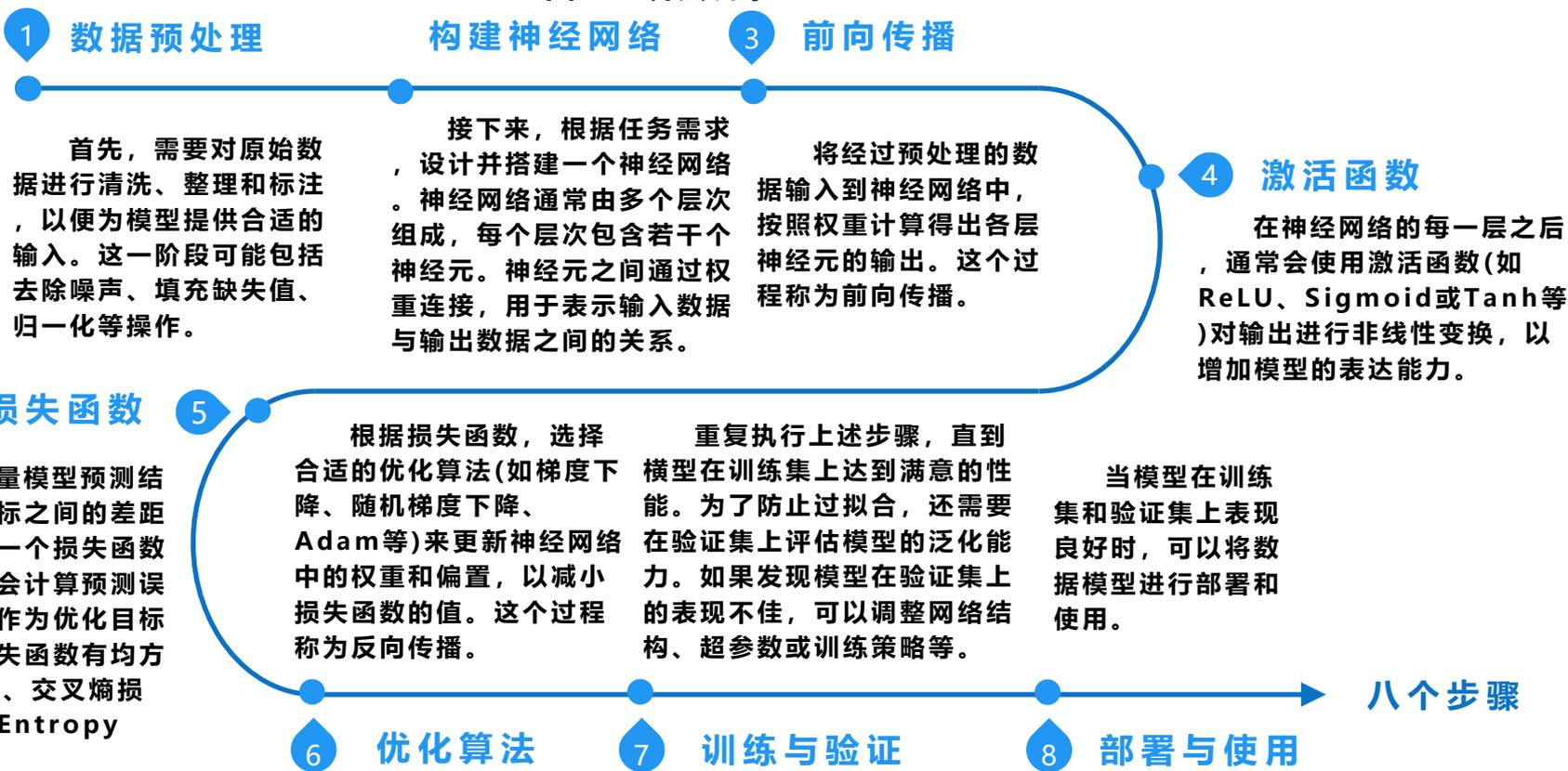
深度学习是基于深度神经网络的，而神经网络算法是机器学习模型的一个分支，包括卷积神经网络 CNN/循环神经网络 RNN 等等，自注意力机制 (Transformer) 则是基于全连接神经网络和循环神经网络的衍生。

深度学习使用多层神经网络，从原始输入中逐步提取更高层次更抽象的特征用于后续算法识别，处理大规模数据是其核心优势。当前，深度学习已经应用到包括图像识别、自然语言处理、语音识别等各领域。



AI大模型是指具有超大规模参数（通常在十亿个以上）、超强计算资源的机器学习模型，其目标是通过增加模型的参数数量来提高模型的表现能力，它们能够处理海量数据，完成各种复杂任务。AI大模型的原理是基于神经网络和大量数据的训练，模型通过模拟人脑的神经元结构，对输入数据进行多层抽象和处理，从而实现了对复杂任务的学习和预测。AI大模型的训练主要分为：**数据预处理、模型构建、模型训练、模型评估**等几大步骤，如下：

图：AI 算法训练



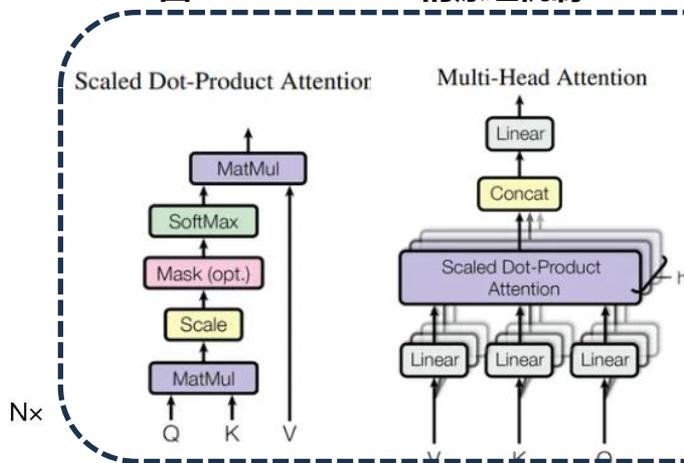
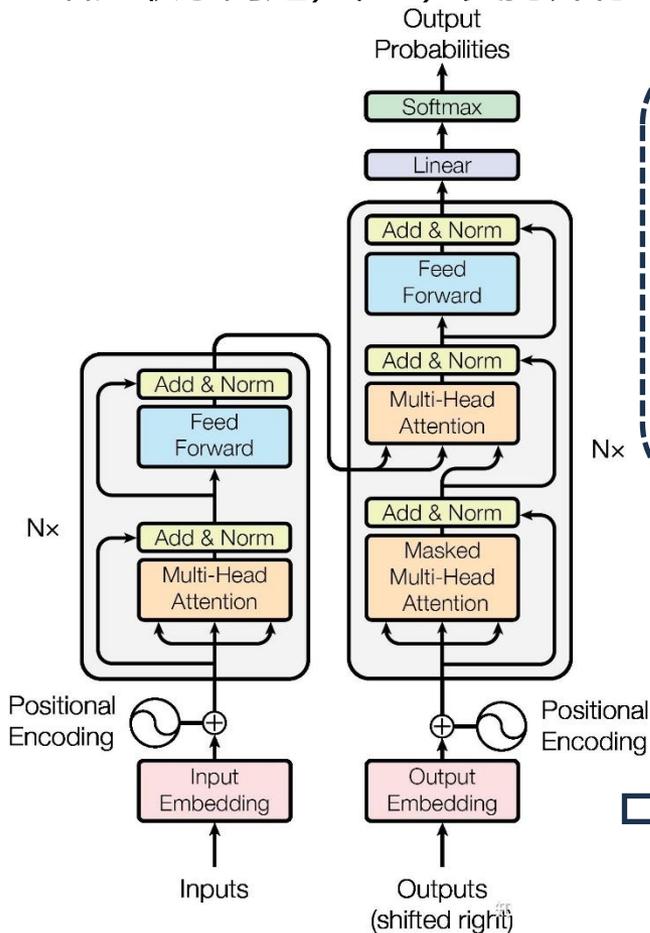
AI大模型能够处理以下几类核心问题：**1) 自然语言处理**：以GPT-3和BERT为例，AI大模型通过学习海量的语料库和上下文，让计算机更加准确地理解和处理自然语言，如翻译、问答、分词、文本生成等领域。**2) 计算机视觉**：以ResNet和EficientNet为例，AI大模型通过学习大量的图像数据和构建更深更复杂的神经网络，使计算机能够对图像进行更加准确的识别和分析，包括目标检测、图像分类、语义分割等领域。**3) 语音识别和生成**。通过以上几类问题的解决，AI大模型可以进一步通过自动化和智能化的方式提高生产效率，在部分工业领域可以实现人机合作或自动化，减少人力成本。

图：AI大模型的优劣势分析



- **Transformer**是本轮大模型颠覆全行业算法架构的核心，也是多模态模型相比之前单一文本理解模型提升的关键，Transformer赋予算法更精准的特征提取能力，强化理解和识别功能，其独特的**自注意力机制**是灵魂，即**Attention is all you need**。
- **Transformer的优势**在于：1) 自注意力机制赋予的长依赖语义问题（捕捉间隔较远的词之间的语义联系问题）；2) 支持并行计算，可极大的提升大模型数据处理效率。

图：Transformer的原理机制



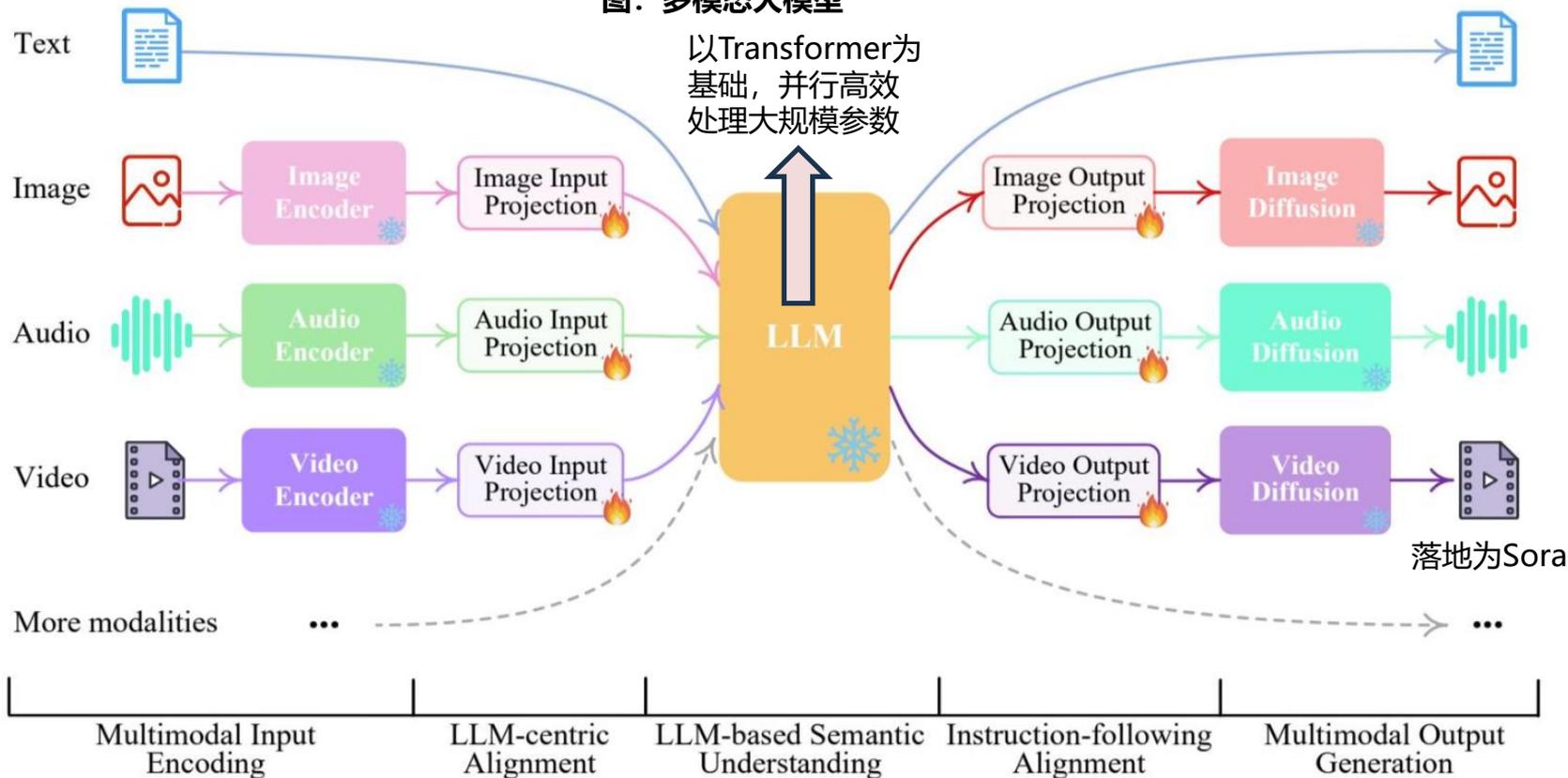
自注意力机制：查询/定位，赋予权重（即注意力），多头即多个维度计算相似度特征，保证充分性。

优势：自注意力机制不同于RNN，无主次先后，可支持并行计算，同时利用位置编码和权重赋予进行定位和筛选。

- input embedding是编码器的输入，
- output embedding是解码器的输入
- Positional Encoding代表位置编码，记住顺序
- Encoder和Decoder分别有6层 (blocks)，Encoder负责特征提取，Decoder则利用特征来完成识别、分类、回归等任务
- 解码后输出结果

■ **由理解内容至生成内容，多模态大语言模型持续进化。**多模态模型是一种能够处理多种类型数据（如文本、图像、音频和视频）的人工智能模型。这种模型的目标是通过结合不同类型的数据来提供更全面、更准确的信息。在自然语言处理（NLP）领域，多模态模型可以用于机器翻译、情感分析、文本摘要等任务。在计算机视觉领域，多模态模型可以用于图像分类、目标检测、人脸识别等任务。**多模态大语言（MM-LLMs）**即是将多模态模型与具备强大推理和生成能力的大语言模型结合的产物，其难点在于如何对齐本不兼容的图像/视频/文本等的编码器。

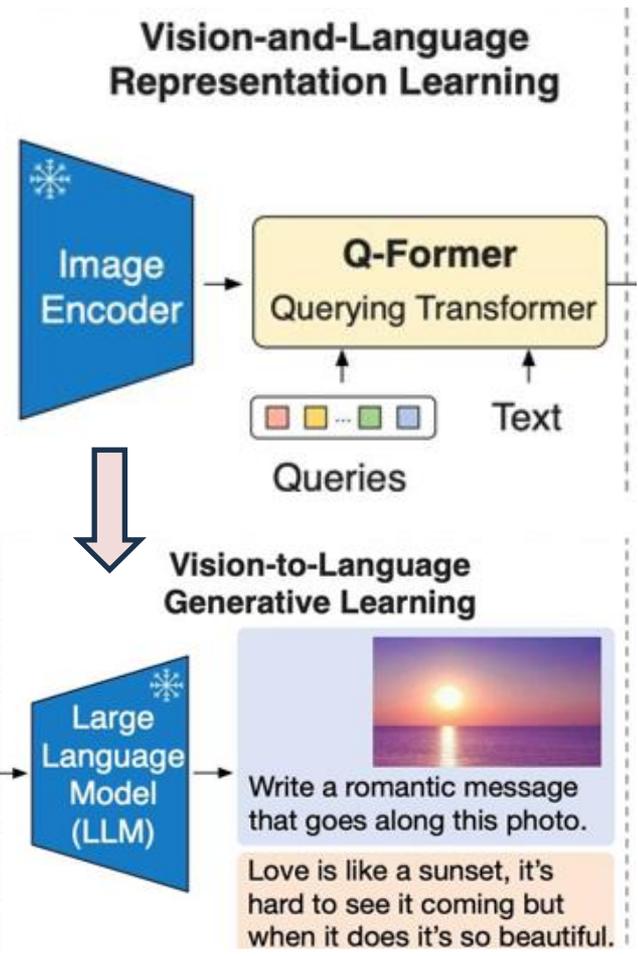
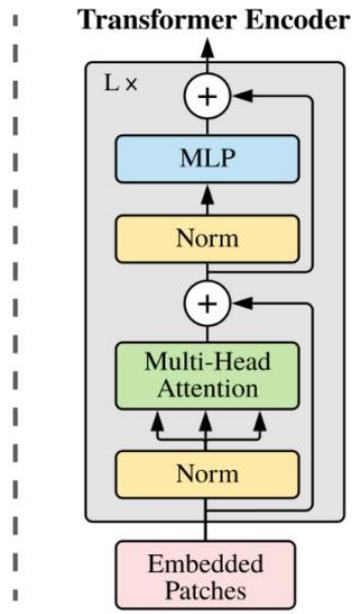
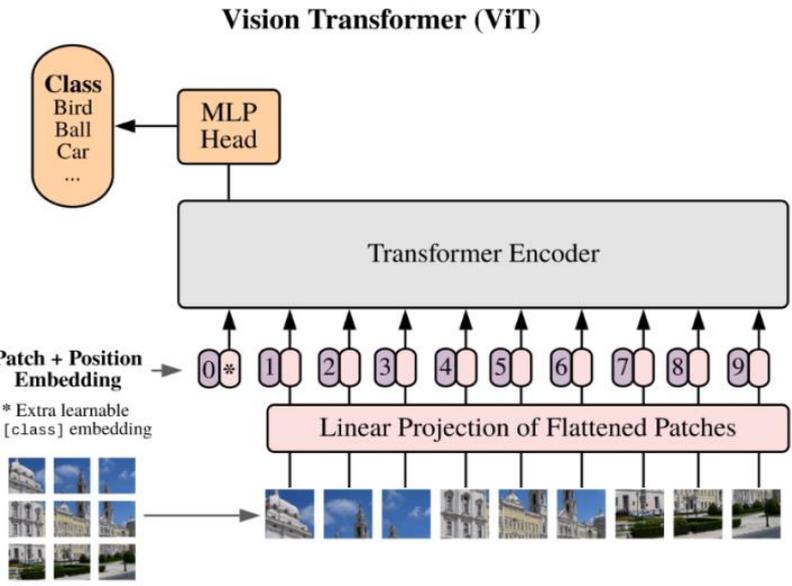
图：多模态大模型



■ **视觉表征主框架由CNN切换Transformer，即ViT**，其**多头自注意力机制**赋予模型**并行高效计算**以及**把握前后长时间依赖关系**的能力，能够同时接收来自文本/图像/音频/视频的特征Token，并接入全链接层服务于下游的分类任务。**ViT成功的秘诀在于大量的数据做预训练**，如果没有这个过程，在开源任务上直接训练，其效果仍会逊色于具有更强归纳偏置的CNN网络。

■ ViT步骤分为三大步：

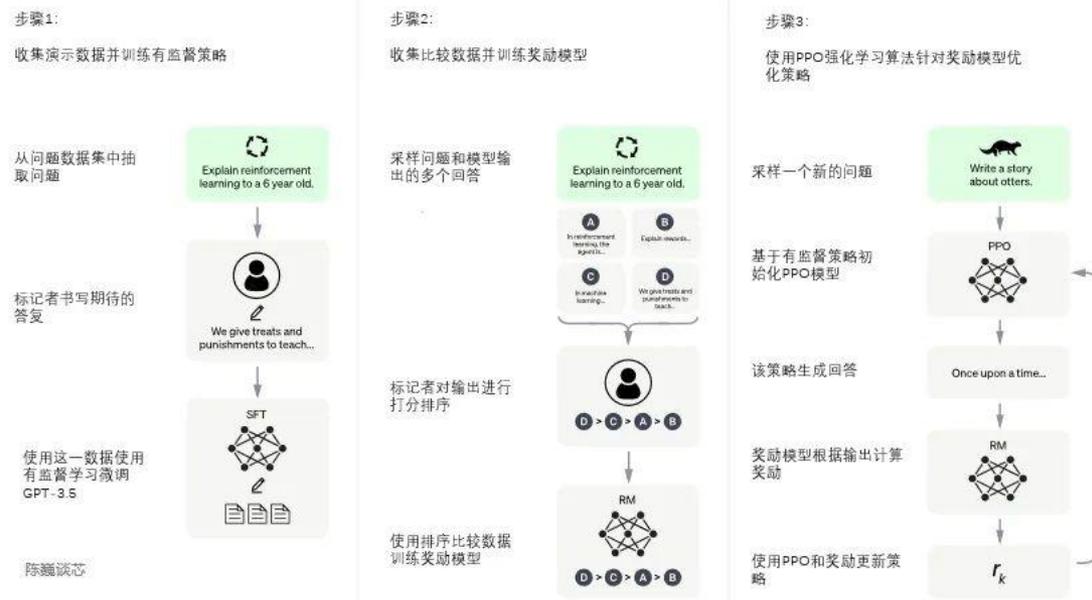
- 图形切块 Patch Embedding；位置编码 Position Embedding；
- **特征提取** Class Token；**注意力权重**赋予 Transformer Encoder；
- 多头输出 MLP Head。



- **GPT1**: 用Transformer的解码器和大量的无标签样本去预训练一个语言模型，然后在子任务上提供少量的标注样本做微调，就可以很大的提高模型的性能。
- **GPT2**: **Zero-shot**，在子任务上不去提供任何相关的训练样本，而是直接用足够大的预训练模型去理解自然语言表达的要求，并基于此做预测。但GPT2性能差，有效性低。
- **GPT3**: **few-shot learning**，兼顾少样本和有效性。用有限的样本，模型可以迅速学会任务。
- **GPT4**: GPT1~3本质还是通过海量的参数学习海量的数据，然后依赖transformer强大的拟合能力使得模型能够收敛，因此不具备文本生成能力。ChatGPT则在此基础上，依赖**指令学习 (Instruction Learning)** 和**人工反馈强化学习 (RLHF)** 进行训练，对原有模型进行有监督的微调（人工优化数据集）+强化学习对模型生成结果进行打分，提高泛化能力，在此基础上loss持续迭代，生成正确结果。相比GPT3规模大幅提升，从1750亿提升至1.8万亿，算力需求大幅提升。

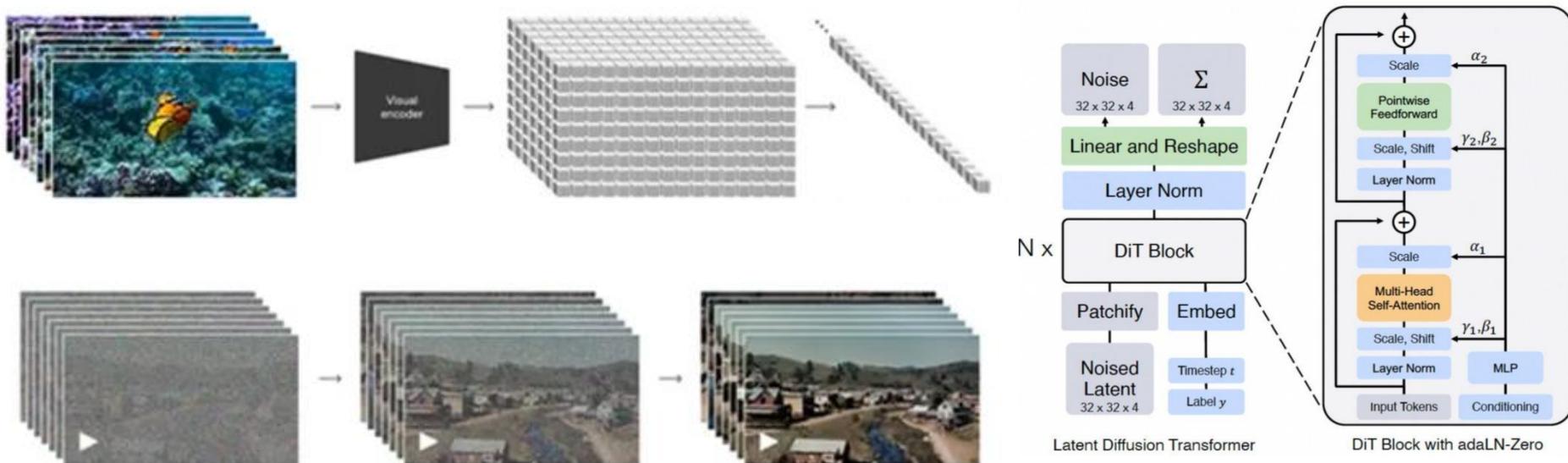
■ **GPT4o**: **完全统一多模态**。o即 omni，意为“全体”、“所有”或“全面的”，打通多模态间输入交互，延迟降低并完全开放免费，并进一步增强推理能力。但其依然是基于Transformer架构去实现对于模态信息的理解和生成，因此并未有底层架构的创新。

图：ChatGPT模型的训练过程



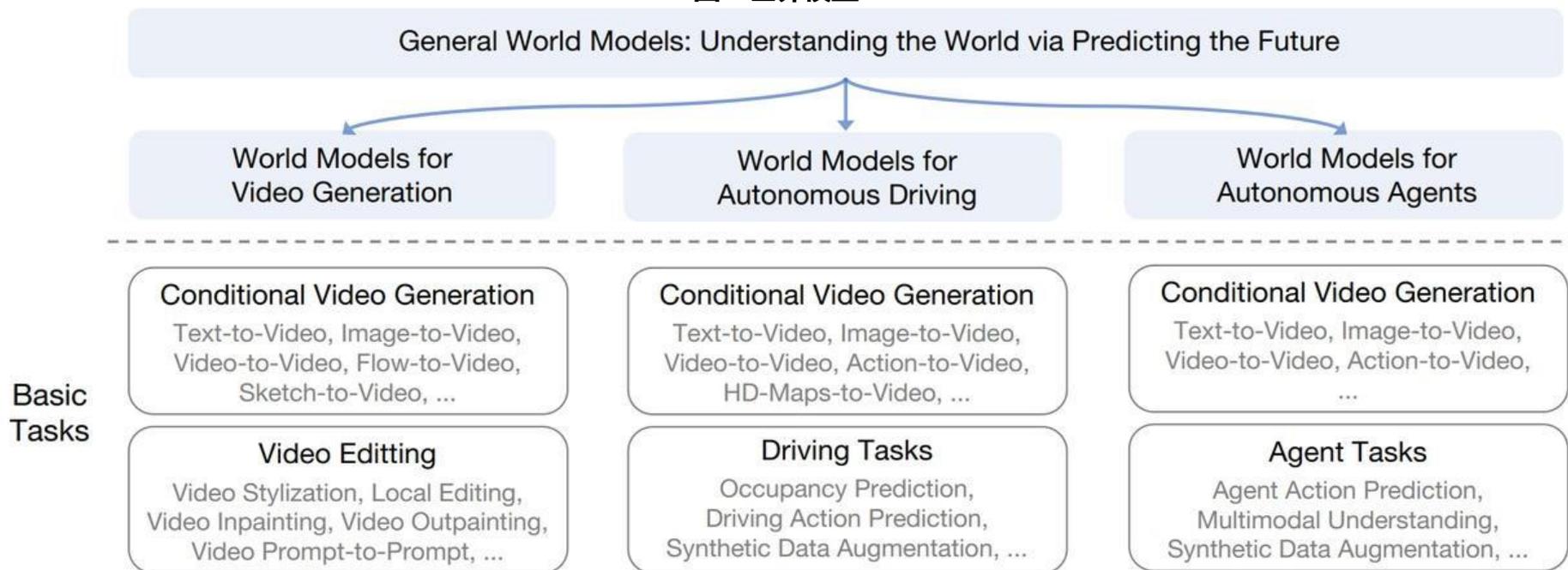
- Sora模型充分利用**扩散模型 (Diffusion Model)** 精细化生成能力以及**Transformer**的前后文全局关系捕捉能力，实现视频每一帧的图像精确生成以及前后的时空一致性。
- Sora可以理解为是**Open AI大模型技术的完全集成**，其原理可以分为三步：1) 首先，模型将视频压缩到**低维潜在空间**中，然后将其分解为patch（类似于GPT中的Token），从而将视频完全压缩。2) 其次，视频patch在低维空间中训练，**扩散模型**通过在训练数据上逐步添加高斯噪声并学习如何逆向去除噪声的过程来生成新数据，并整合了Transformer的多头注意力机制和自适应层归一化，提升扩散模型在处理连续大规模视频数据时的**稳定性和一致性**。3) 最后，模型通过对应解码器，将生成的元素**映射回像素空间**，完成视频生成任务。
- Sora的核心DiT模型：**Transformer架构取代原有UNet架构**（本质为CNN卷积神经网络），提升长时间性能的同时具备可扩展性+灵活性，并支持跨领域扩展，或可应用至自动驾驶领域。

图：Sora文生视频，DiT模型为基础原理



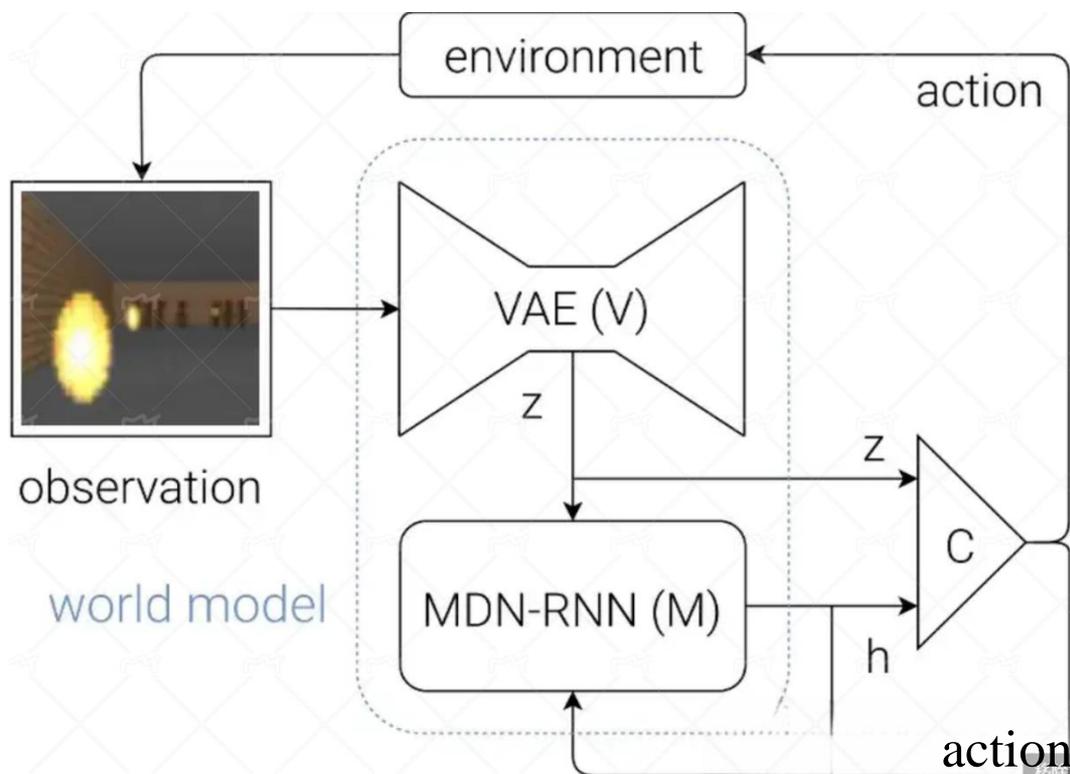
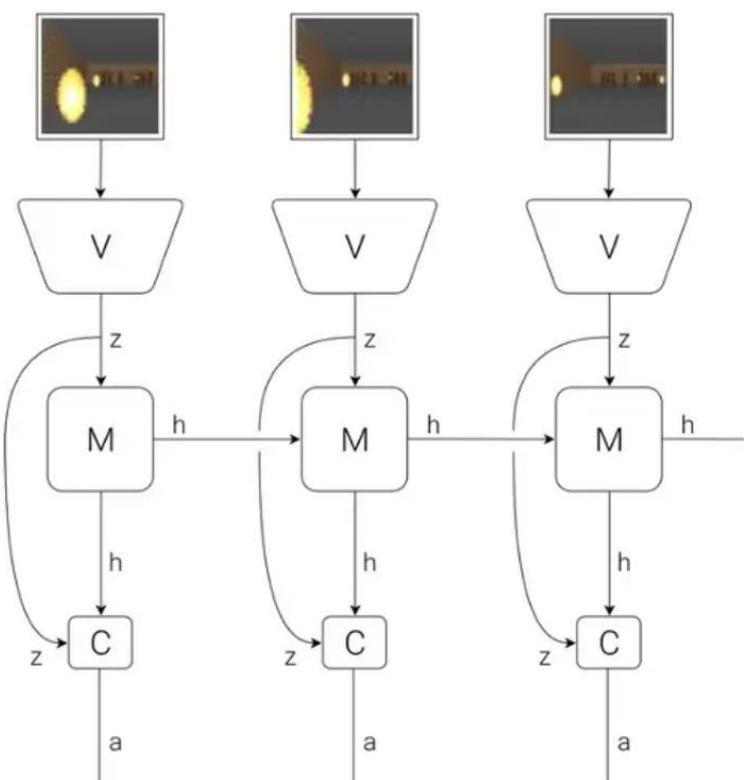
- **视频生成Video Generation是世界模型World Model的基础，Sora的成功使得DiT扩散模型收敛为视频生成的主导方式。** 基于对历史信息的理解生成对于未来的预测进而合理表征，并结合可能的行为action进一步辅助预测可能的结果，可应用于包括直接的视频生成以及自动驾驶和机器人等通用人工智能多个领域。
- **基于数据，超越数据。** 传统大模型要适应新的精细任务时，必须基于数据对模型参数进行全面微调，依赖【预训练】环节，不同任务与不同数据一一对应，模型专业化。相比过往大模型强调的模态信息理解的能力，世界模型更加注重【因果和反事实推理、模拟客观物理定律】等规划和预测的能力，并具备**强泛化和高效率**等性能表现。

图：世界模型



- 传统的多子函数模块化算法演变为世界模型数据闭环训练，持续优化提升算法认知。世界模型在进行端到端训练的过程为：
- 观测值经过V（基于Transformer的Vision Model，图中为原先采用的VAE架构）提取feature，然后经过M（基于LSTM的Memory）得到h（预测值），最后预测值和历史信息合并至C（Controller）得到动作，基于动作和环境交互的结果产生新的观测值，保障车辆做出合理行为。

图：传统模块化算法架构【V+M+C】转变为端到端【VM输出C】



二、车端：大模型重塑智驾算法架构

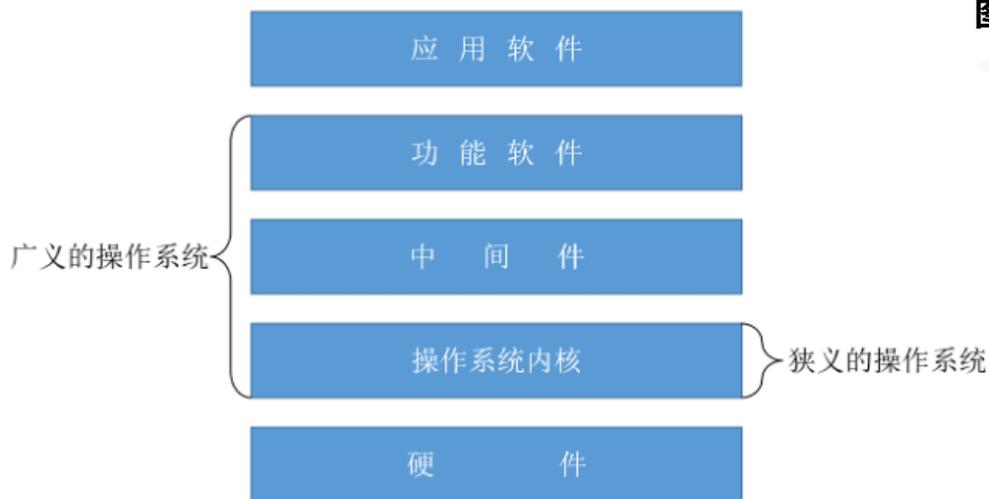
- **底层调动支持：**1) **异构化底层软件配合SoC异构芯片支持不同类型的功能实现落地。** 用于智驾域控的SoC异构芯片是高度集成化的，除多核CPU外，还会集成DSP/NPU或其它专用计算加速单元。在此硬件上进行匹配的异构软件，多核CPU经过虚拟化之后，可以在不同的虚拟机上运行不同的操作系统，例如运行Linux来充分利用Linux社区丰富的软件生态，运行QNX/VxWorks 来达到更好的实时性。2) **SOA中间件提供标准化接口服务。** SOA中间件能屏蔽不同服务间软硬件异构平台的差异性，一方面让异构平台都能够以SOA的方式交互，另一方面让通用的功能逻辑能在不同硬件平台之间移植。比如将DSP或NPU的使用封装在特定的库中，对外呈现为标准SOA服务。3) **软件框架：**即可复用的设计构件，它规定了应用的体系结构，代表已经完成项目的底层开发基础搭建，可以在其基础上差异化开发的半成品。可以让使用者减少很多重复的代码、让代码的结构更加清晰，耦合度更低，后期维护方便。
- **上层应用赋能：算法：**即解决某一类问题的特定策略机制，不同算法解决同一问题时的质量优劣不同，这将影响程序的效率。一个算法的评价主要从时间复杂度和空间复杂度来考虑。



图：汽车智能驾驶软件层级架构

多自研，匹配上层应用算法
多采用第三方成熟系统

- **What:** 传统汽车时代，狭义的操作系统指操作系统内核；当前智电汽车域控时代，系统更加复杂，因此需要基于内核进行大量工程优化以后方可直接开发上层应用软件，中间件和功能软件即由工程优化算法标准化后所得。当前广义操作系统包括内核以及中间件和功能软件等几部分。
- **How:** 高性能、高安全、跨平台、高效开发应用。操作系统要求较强实时性，系统任务调度时钟周期在毫秒级；且提供高可靠性和较强功能安全。未来，行业舱驾一体化等集中式EE架构还要求操作系统保证合理共享算力资源，支持系统快速OTA迭代。
- **底层支持软件相对标准化，通用性较强。** 1) **操作系统内核 (Kernel) :** 即为狭义操作系统，如OSEK OS、VxWorks、RT-Linux等。内核提供操作系统最基本的功能，负责管理系统的进程、内存、设备驱动程序、文件和网络系统，决定着系统的性能和稳定性。2) **中间件:** 处于应用和操作系统之间的软件，实现异构网络环境下软件互联和互操作等共性问题，提供标准接口、协议，具有较高的移植性，如POSIX/ARA（自适应AutoSAR运行时环境即中间件API接口）和DDS（分布式实时通信中间件）。3) **功能软件:** 主要指自动驾驶的核心共性功能模块。

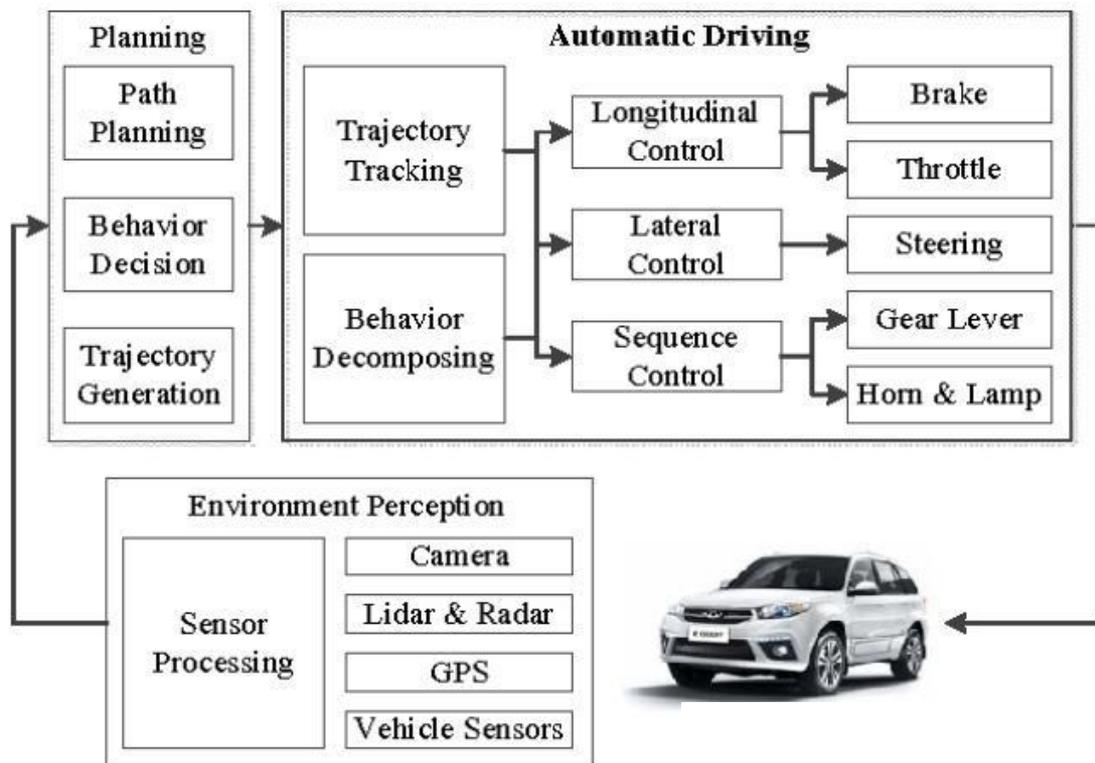


图：汽车操作系统

| 主流操作系统 | 特点 |
|---------|--------------------------------|
| QNX | 可靠性很好，协议栈、各种外设驱动稳定，只是运行所需资源有些多 |
| Linux | 实时性较差，但软件资源内容丰富 |
| VxWorks | 开发工具功能强大，使用方便，但是价格昂贵 |

- 车端上层应用算法分为感知 (Perception) /规控 (Planning and Decision) /执行 (Motion and control) 三大模块。感知模块的输入可以是各种传感器：摄像头、激光雷达、毫米波雷达等，同时感知也会接受车身传感器的信息来对感知结果进行修正。决策规划模型又分为行为决策、轨迹规划、轨迹生成等模块。控制模块主要分为横向控制与纵向控制，横向控制主要是控制转向系统，纵向控制涉及油门与刹车的控制。
- 模块化的软件算法早期具备可解释性强、安全性强、灵活性高的优点，但过于依赖先验数据，无法实时更新，且穷举法终究没办法覆盖所有长尾场景，因此随智驾等级提升，正逐渐被大模型端到端所取代。

图：智能驾驶应用层算法分类

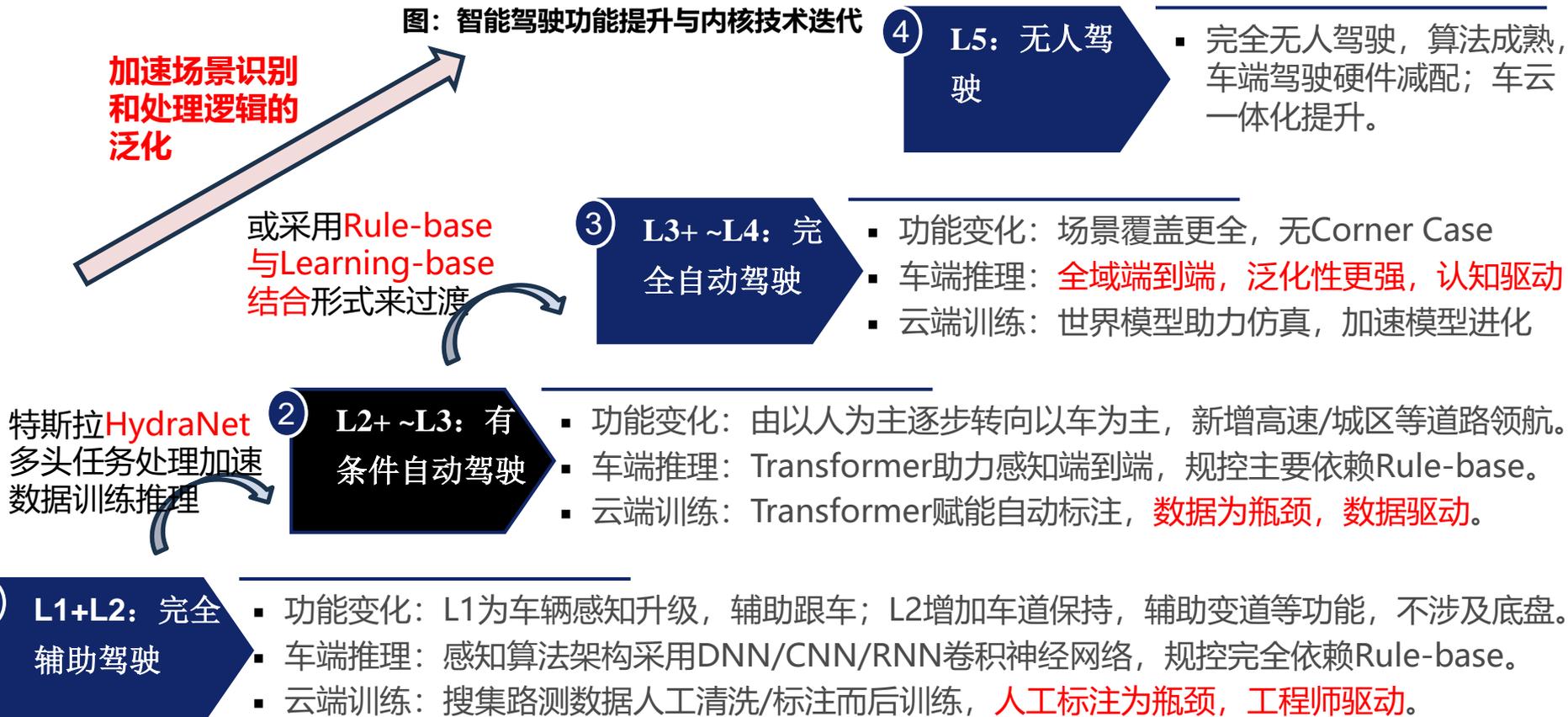


- **多模态数据喂养软件算法，驱动功能进化。** 硬件端升级提供多样化数据，智驾传感器由单目视觉向双目视觉以及视觉+雷达的方向演变，多传感器支持对周围环境的更精确识别，图像/点云等数据呈现方式多样化，数据赋能算法升级，支持智驾功能由L3级别以下向L3及以上迭代升级。
- **场景泛化是智驾能力提升的重要方向，数据需求量激增驱动算法由小模型切换为大模型。** 智驾软件上层应用算法中，感知能力提升核心系**场景识别能力的泛化**，规控能力提升核心系**场景处理逻辑的泛化**，均以**有效数据量为核心驱动**，数据处理/利用等是瓶颈。DNN/CNN/RNN小模型切换为Transformer大模型赋能**多模态数据处理**，重塑车端算法架构，以支持更高级别智驾功能。
- **世界模型（自学习，强泛化，高效率）或为通向L4级完全自动驾驶的必由之路。** 数据赋能催化模型自学习，理解物理世界基础运行规律，模型“认知能力”提升，或可实现完全自动驾驶。

| 类型 | 算法架构 | 特征 |
|-----|------------------------|---|
| 小模型 | DNN | 通过对像素级别进行逐个对比，通过图像特征进行识别，精确，存储简单，但计算量较大 |
| | CNN | 以“卷积-池化”的方式，将整幅图按兴趣区域 ROI 进行特征分割后，分别对 ROI 进行局部像素对比识别，相当于在DNN基础上进行降维，特征提取降低计算量需求 |
| | RNN | 在CNN基础上引入循环核，便于处理时间序列图像特征 |
| 大模型 | Transformer | 基于特有的 自注意力机制 ，能够有效捕捉序列信息中 长距离依赖关系并支持并行计算 ，适合处理大量且多类型的数据 |
| | ResNet | 深度 卷积神经网络，为解决随网络层数增加带来的网络退化（梯度消失和梯度爆炸）现象，残差网络使用跳跃连接实现信号跨层传播 |
| | VQ-VAE | 以自编码器将图像压缩为有限数量的向量集合，从而实现了高效的图像表示和重构 |
| | 生成对抗网络GAN | GAN由生成器和判别器组成，生成器经由输入的噪声向量输出为假设图像，判别器区分假设图像和真实图像并反馈训练，驱动生成更加逼真的图像 |
| | 世界模型（基于Transformer的复合） | 人工智能系统构建的对于现实世界的模拟与表达，模型充分理解物理世界的客观运行规律 |

■ 产业对于智驾功能持续迭代的诉求驱动车端/云端算法持续升级。1) 阶段一：L3以下。OEM搜集数据清洗标注后训练，算法基于特定规则，场景泛化能力较差，数据以及标注工程师均为瓶颈。2) 阶段二：L3级别。Transformer赋能，大模型端到端算法落地，自动化标注多模态数据，并快速提升数据利用效率，数据与算法能力形成正循环，**场景识别的泛化能力提升**。3) 阶段三：L3以上。车端算法完全端到端，Learning-base主导算法训练，强化算法感知以及规控端的泛化能力，**算法认知规划能力提升**，逐步支持完全自动驾驶。

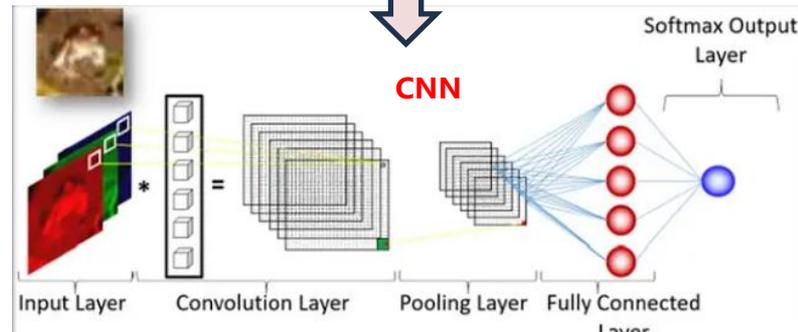
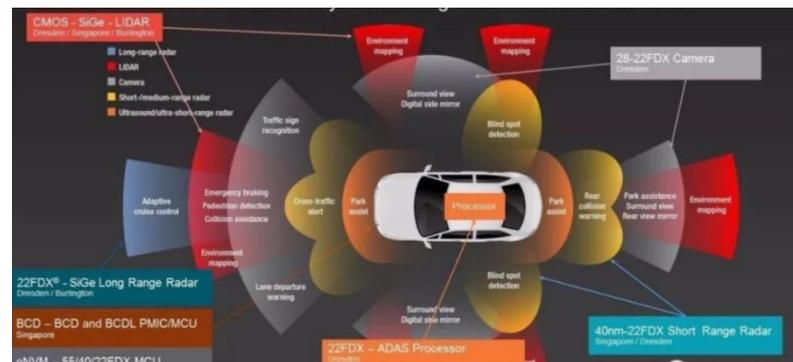
图：智能驾驶功能提升与内核技术迭代



■ **精确感知为跃升L3的瓶颈, BEV+Transformer架构上车支持产业化解决方案落地。** L1-L2: 智驾迭代主要集中在感知维度的升级, 不包含规控以及执行端底盘的变化。由L2-L3的过程, 则是由以人为主体逐步切换为人车共驾, 【外部环境的精确感知问题】为瓶颈。当前, 低成本的软件算法升级 (BEV+Transformer) 基本取代高成本解决方案 (激光雷达+高精地图), 推动产业化。

➤ **传统CV小模型架构下, 感知能力升级更依赖硬件。** 依赖摄像头+激光雷达/毫米波雷达等能够感知三维深度的硬件实时描绘周围环境, 并配合高精度地图提供的先验信息辅助判断; 算法领域, 以CNN卷积神经网络进行图像的特征提取和目标识别, RNN大规模记忆处理时间序列数据。

图: 智能驾驶感知



■ 暴露问题:

- ✓ 能力提升依赖庞大数据量, 但数据量过大带来: 1) 算法难以捕捉长距离依赖关系; 2) 数据处理依赖人工, 利用效率较低。
- ✓ 雷达硬件成本过高且高精度地图覆盖面较窄。

精确感知为跃升L3的瓶颈, BEV+Transformer架构上车支持产业化解决方案落地。

- **BEV+Transformer大模型架构赋能精准感知【处理长距离依赖关系+并行计算+统一融合】。**
- ✓ **Transformer以自注意力机制完成时间/空间关联, 处理长序列数据信息同时支持并行计算, 提高效率。** 1) 高效对多模态数据进行特征提取, 利用空间元素间的关联关系进行全局建模, 实现特征增强; 2) 利用元素与前后体之间的关联关系, 高效处理长序列信息, 避免重复递归/卷积。
- ✓ **BEVFusion统一框架, 实现多模态信息的精确端到端融合。** 用独立流程分别处理雷达和摄像头数据, 然后在鸟瞰视图(BEV)层面进行融合, 并扩展到多任务、多传感器框架, 兼顾信息完整度和效率, 标准化框架也便于数据清洗标注。

图: Self-Attention (自注意力机制) 原理

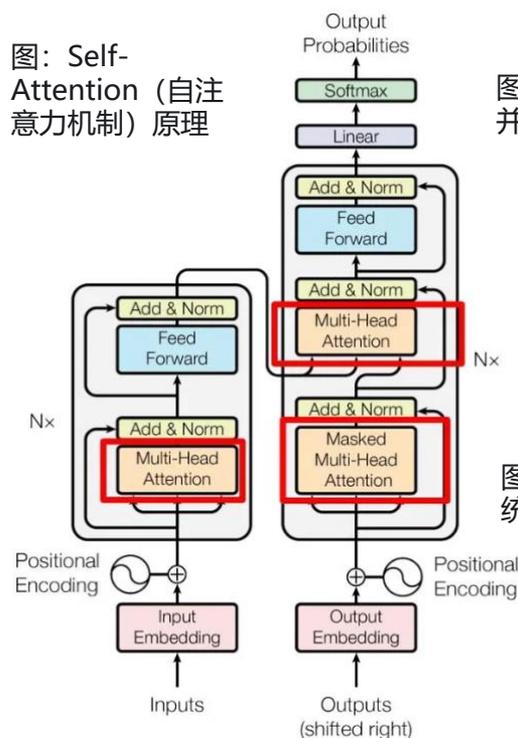


图: Transformer 并行特征提取

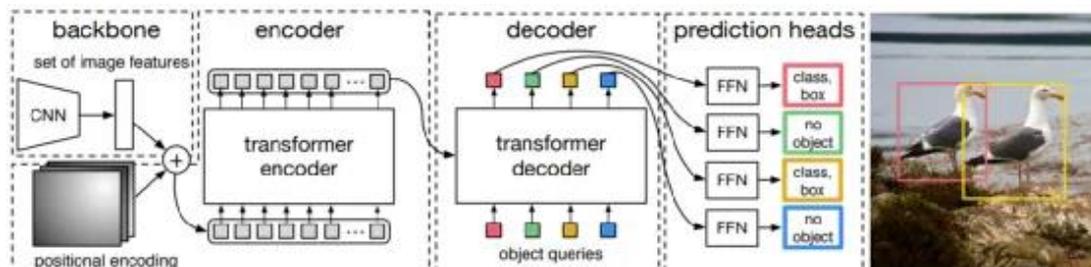
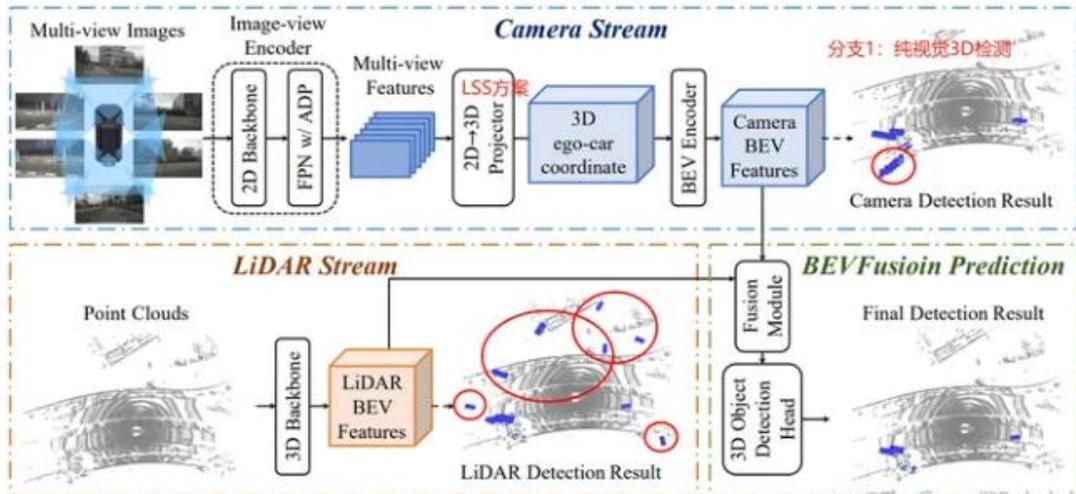


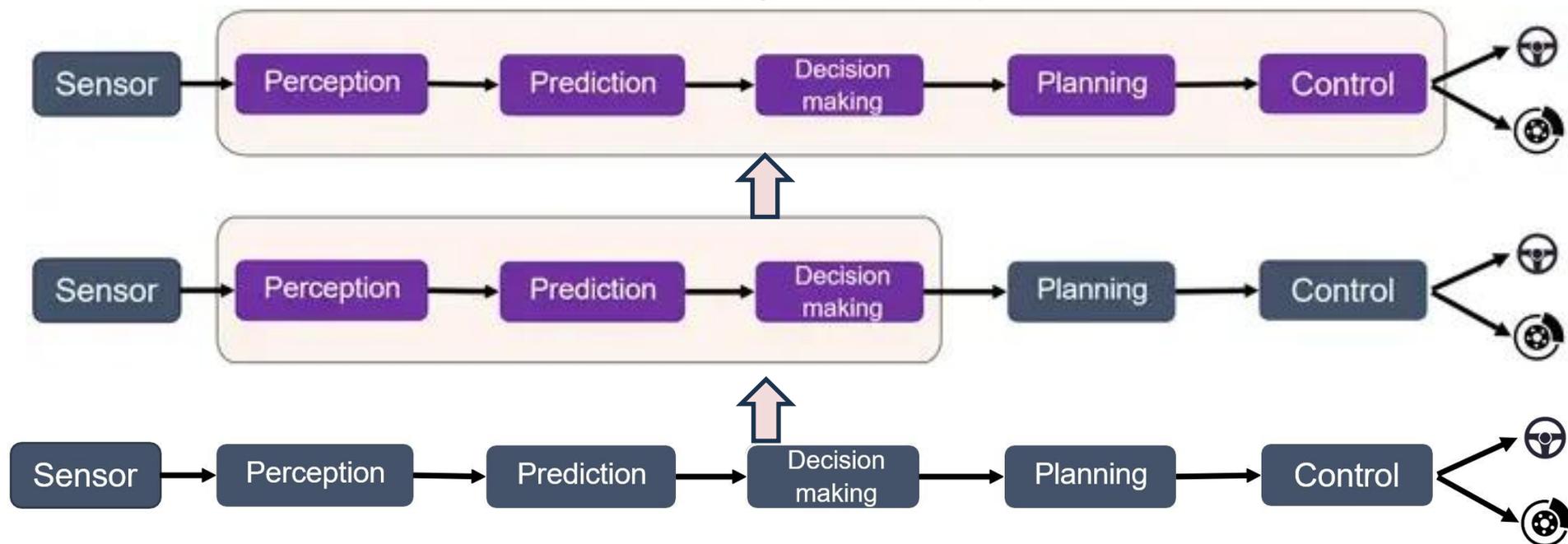
图: BEV视图进行统一特征级融合



■ **全面泛化**为跃升L4的瓶颈, World Model世界模型 (本质为模型基于对客观世界物理规律的理解进行精准预测) 或为终局解决方案。L4智驾要求完全自动驾驶, 相比L3智驾, 其对车辆运行过程中的场景识别以及预测规划提出更高要求, 强调【100%全面泛化】。因此, 车辆自主控制阶段的算法瓶颈转变为【如何正确预测周围环境变化并合理规划行驶路径】。

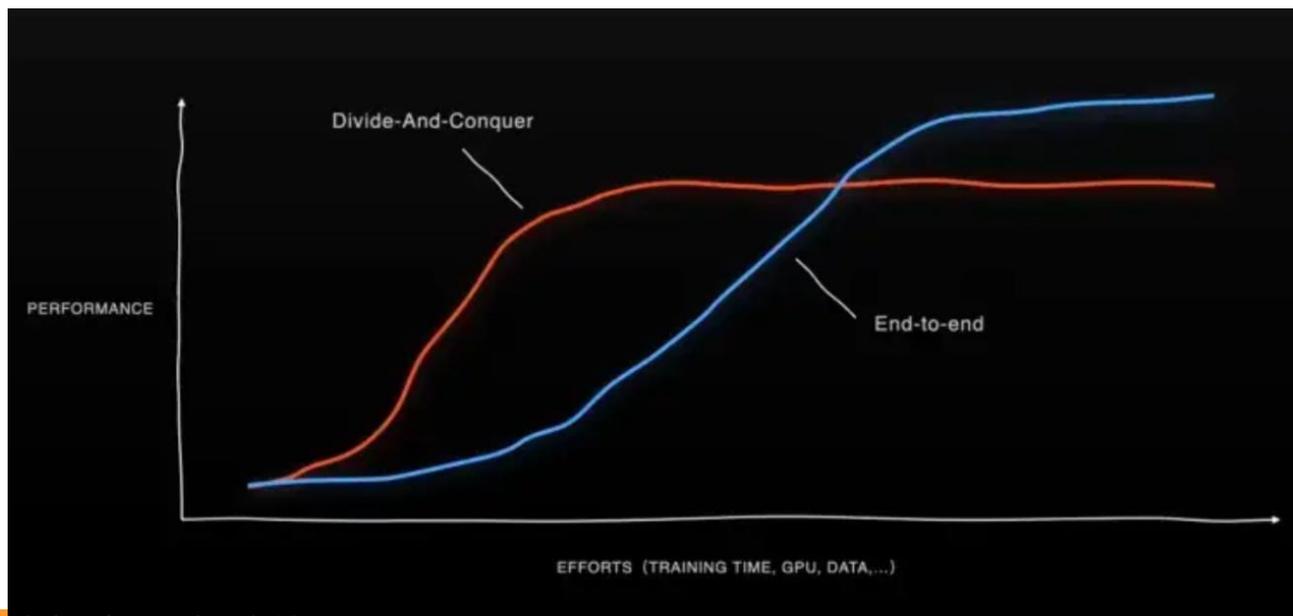
➤ **世界模型或为L4完全自动驾驶终局解决方案**, 通过**End-to-End算法**+大数据闭环训练, 培养算法认知泛化能力, 实现【通用具身智能】。World Model的核心增量: **反事实推理** (Counterfactual reasoning), 也即对于数据中没有见过的决策, 在world model中都能推理出决策的结果。

图: 模块化算法架构逐步迭代为端到端



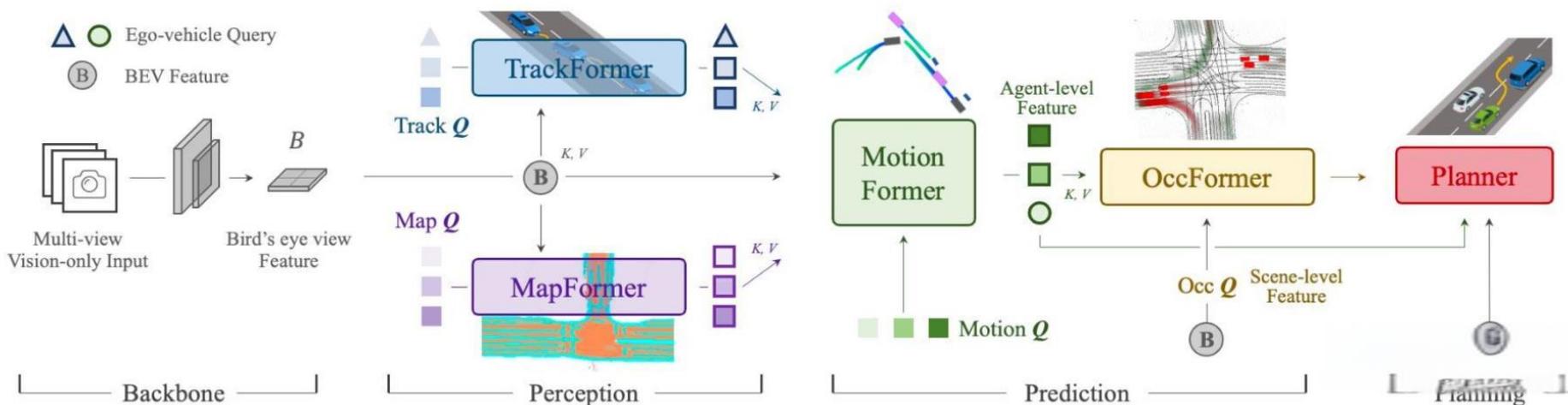
■ 端到端架构将感知-规控-执行模块串联统一训练。

- **优点：**1) **结构简单直接，算法工作量较低，减少维护成本；**传统策略基于完整规则进行控制，行为机械且代码量大，据小鹏2022年，城市NGP代码量是高速的88倍，端到端有明显优势，架构和代码方面一劳永逸，高度整合，省去独立模块开发以及数据人工标注和规则设计，降低成本。2) **更利于追求“全局最优解”，能力上限高，泛化能力更强，完美应对长尾场景；**基于规则的方案遇到未学习过的场景时无法做出良好的处理策略，且基于已有数据难以保证100%的场景数据覆盖，传统模式的最优解理论上无法达成，端到端是通过对场景的理解进行判断，天花板更高。3) **减少数据时延，便于车端迅速判断，提升安全系数。**
- **缺点：**数据的针对性更弱、模型训练的信号更弱，意味着提升性能**所需的数据量和算力规模更大。**端到端的优势在数据量达到一定程度后性能显著提高，但数据量较小时性能上升缓慢，远低于解耦的传统基于专家模型的策略（特斯拉当前高速依然保留FSDv11的代码）。



- **UniAD (感知Perception+规划Planning)**：多个共享BEV 特征的Transformer网络将跟踪+建图+轨迹预测+占据栅格预测统一到一起，并且使用不依赖高精地图的Planner 作为一个最终的目标输出，同时使用Plan结果作为整体训练的loss 来源。相比于特斯拉FSDv12的黑盒端到端，其在感知/规控等模块间虽然一体化整合，但并非一个整体黑盒网络，仍可以对各个模块进行分别的监测和优化，各个模块间有了相当的可解释性，也有利于训练和Debug，【解耦白盒】是核心。
- **UniSim (仿真)**：使用多个NeRF 网络，重建静态背景，并且将每个动态物体也构建出来，同时保存动态物体的轨迹。之后在仿真环境中，静态物体作为整体世界背景，所有的动态物体根据轨迹转换到世界坐标系中，进行世界环境的模拟。利用动静态解耦的方式，为之后仿真物体的删除和插入提供便利，同时对危险场景的模拟提供充足的自由度，保证场景泛化。

图：UniAD算法框架



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/015111301043011222>