

OpenAI Sora专题

Transformer扩展优势凸显，视频理解与
生成能力提升

西南证券研究发展中心
海外研究团队 王湘杰
2024年2月

核心观点

- **事件**：2024年2月16日，OpenAI发布文生视频模型——Sora及其技术报告《Video generation models as world simulators》。
- **从Sora模型看文生视频的技术路径**：技术路径尚未收敛，Transformer扩展特性优势凸显。市场大多认为扩散模型是图像和视频生成领域的主流路径，但没有重视Transformer架构scale up的能力。OpenAI技术报告指出，Sora是基于扩散模型，但更强调，Sora是一个基于Transformer架构的扩散模型，其优秀的生成能力离不开Transformer架构优秀的scaling特性。当前，为构建性能更优、效率更高的视频生成模型，已出现多种结合Diffusion Model和Transformer架构的构建方式。
- **从Sora模型看文生视频的最新能力**（假设展示视频可以代表Sora的一般性表现）：**1）强大的理解能力**：Sora模型不仅可以理解Prompt的内容，还能理解事物在物理世界中的存在方式，突出的语言理解能力是其能够准确生成视频的前提。**2）优秀的生成能力**：①长度：可生成60s视频；②复杂度：能够生成包含多个角色、多种主题、多类运动形态的复杂场景；③逼真度：能够呈现更多的视觉细节，具备更高清的画质，为用户提供逼真的视觉体验；④连贯性&一致性：可以生成同一角色的多个镜头，保持其在整个视频中的外观，在角度移动和切换情况下，人物和场景元素在三维空间中的位置关系能够保持一致的运动；⑤可控性：在某一Prompt基础上只改动一个关键词仍能生成优质的视频，具备较好的可控性。**3）其他能力**：图生视频，视频扩展/编辑/拼接/模拟等。
- **从Sora模型看文生视频的行业影响**：目前OpenAI已向部分视觉艺术家、设计师和电影制作人提供访问权限，以获取专业的反馈。我们认为，以Sora为代表的视频生成模型有望给广告/设计/短视频/游戏等行业带来变化。从中短期来看，视频生成模型将更多的作为创作工具对相关行业进行赋能。
- **投资建议**：随着文生视频模型的迭代升级和未来的大规模应用，算力需求有望增长，云服务有望成为算力的重要补充，视频传输也将带动高带宽、高性能的光通信需求，建议关注算力、云服务和光通信领域投资机会。相关标的：英伟达(NVDA.O)、超威半导体(AMD.O)、微软(MSFT.O)等。
- **风险提示**：技术进展不及预期风险；行业竞争加剧风险；应用开发不及预期风险。

目 录

1 从Sora看文生视频模型的技术路径

1.1 发展现状：视觉生成热潮兴起，路径探索仍在继续

1.2 路径对比：扩散模型占据主流，自回归模型潜力可期

1.3 未来趋势：transformer优势凸显，多种模型构建方式有望呈现

2 从Sora看文生视频模型的最新能力

2.1 模型背景：团队成员年轻有为，相关技术积累深厚

2.2 理解能力：领会并细化提示词，学习及模拟物理世界

2.3 生成能力：视频长度优势明显，连贯性及可控性提升

2.4 其他能力：图生视频，视频扩展/编辑/拼接/模拟等

3 从Sora看文生视频模型的影响

3.1 算力端：参数扩展提升训练算力消耗，应用落地拉动推理算力增长

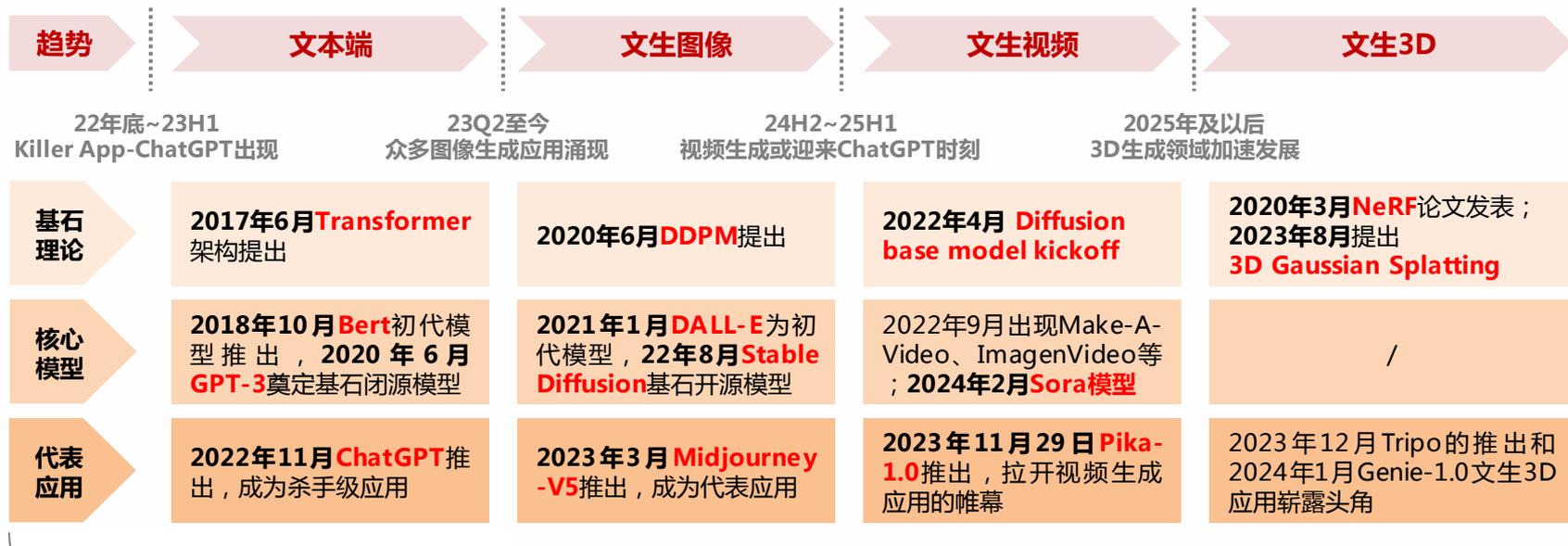
3.2 应用端：文生视频模型拓宽应用广度，前期有望以助手角色落地

3.3 其他：云服务将成算力重要补充，多媒体传输带动光通信需求

1.1.1 发展现状：多模态发展已成共识，视觉生成热潮逐步兴起

- **从模态发展情况来看**：AI大模型的生成和理解能力从单模态向多模态发展已成社会共识。当前，大语言模型在文本端的应用趋于成熟，图像和视频生成技术快速发展，AI视觉浪潮正加速到来。
- **从市场催化节奏来看**：受文本端杀手级应用ChatGPT（于22年11月推出）和图像生成代表应用Midjourney V5（于23年3月推出）的影响，文本端和图像生成应用已于2023年迎来市场催化。**2024年2月16日，OpenAI发布Sora文生视频，引发市场对视频生成领域的高度关注。**未来，随着模型的持续迭代和相关应用的涌现，**视频生成领域有望在24年迎来热潮。**

大模型多模态发展情况



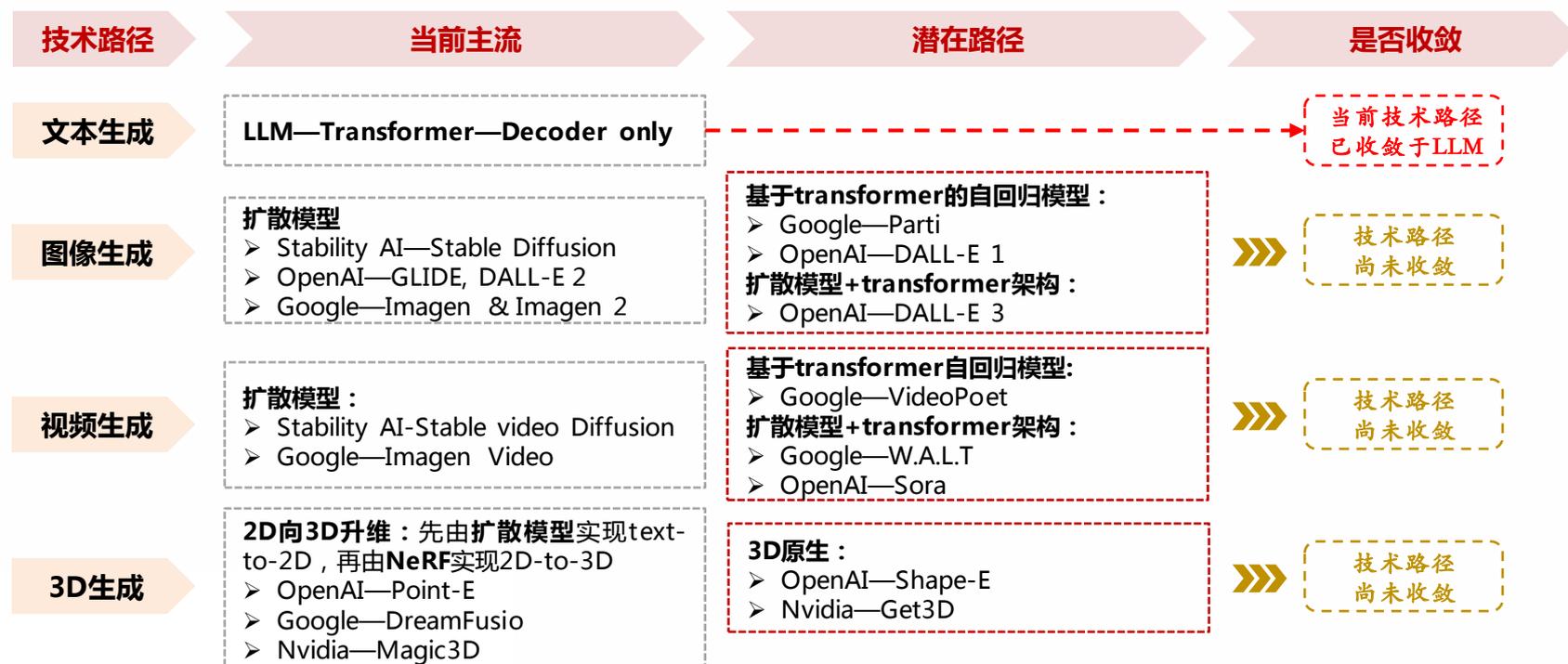
从理论到应用的历时逐步缩短，多模态时代加速到来

资料来源：西南证券

1.1.2 发展现状：文本端技术路径收敛于LLM，视觉生成领域仍在探索

- 从技术路径收敛情况来看：1) 文本生成方面，在OpenAI GPT系列模型的引领下，大语言模型（LLM—large language model）已成为当前文本端的确定性路径。2) 多模态方面，图像和视频生成的技术路径均未收敛。图像和视频生成主要包括两大路径：扩散模型（Diffusion Model）和自回归模型（Autoregressive Model），其中，扩散模型是当前主流，而基于transformer架构的自回归模型具备较大潜力，扩散模型也逐步呈现结合Transformer的发展趋势。

多模态大模型技术路径情况



资料来源：西南证券

1.2.1 路径对比：扩散模型占据主流，自回归模型潜力可期

- ❑ **扩散模型**：成为文生图和文生视频领域**主流技术路径**的核心原因 → **Available**：基于扩散的**Stable Diffusion模型选择开源**，使广大开发者能够在此基础上运用和优化扩散模型，**其公开可用的特性**使之成为图像和视频生成领域的**主流基础模型**；
- ❑ **自回归模型**：在视觉生成领域**具备较大发展潜力**的主要原因 → **Scalable**：基于transformer的自回归模型相较于扩散模型更适合、更容易做规模扩展（scale up）。目前，扩散模型的参数规模通常在billion（十亿）级别，而**自回归模型得益于transformer架构，参数具备更大的扩展潜力**，图像和视频领域的自回归模型有望借鉴transformer在文本领域LLM的经验，通过对不同模态进行跨模态、规模化的建模，实现“大力出奇迹”。

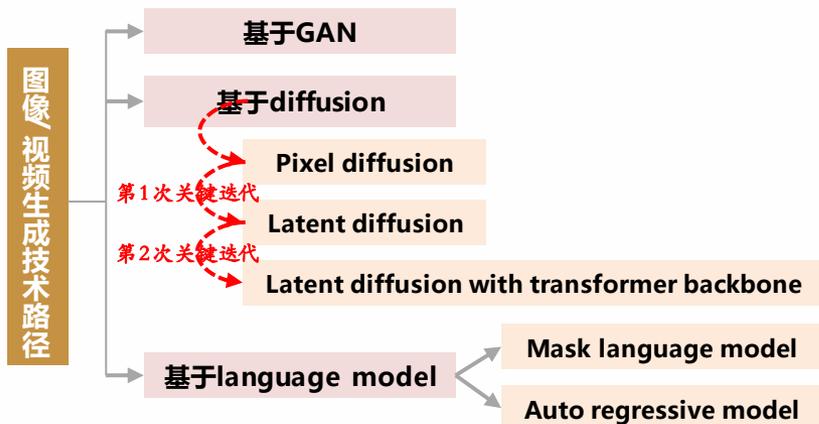
扩散模型 VS 自回归模型

	扩散模型	自回归模型
原理	扩散过程是指对数据逐渐增加高斯噪音直至数据变成随机噪音的过程，扩散模型包含前向扩散过程和反向生成过程。	通过编码器将文本转化成token或序列，应用自回归预测，经过训练好的模型解码输出图像。
优点	①需要更少的计算资源，计算效率更高，成本更低；②更好的缩放，在处理数据时（压缩/放大）能获得高质量图像。	①在规模持续增长的情况下能够呈现更好的生成效果；②视频的时间序列结构适合转化为预测下一帧的任务形态。
缺点	蕴含多个（原始模型可能要上千个）前向传播过程，采样程序效率低下；似然估计欠佳，数据泛化能力差。	自回归模型参数量大，对训练数据的质量和体量要求高，整体训练成本高。
模型	图像：① Stability AI—Stable Diffusion；② OpenAI—GLIDE, DALL-E2；③ Google—Imagen & Imagen 2 视频：① Stability AI—Stable video Diffusion；② Google—Imagen Video；	图像：① Google—Parti；② OpenAI—ImageGPT, DALL-E 视频：① Google VideoPoet
应用	图像：Midjourney； 视频：Runway；Pika 1.0	/

资料来源：西南证券

1.2.1 路径对比：扩散模型占据主流，自回归模型潜力可期

□ 从图像生成模型发展顺序来看：GAN出现最早，其次是基于language model的图像生成模型（2020年6月OpenAI提出ImageGPT），最后是基于diffusion技术的扩散模型。OpenAI于2021年5月提出在扩散过程中使用显式分类器进行引导，展示出扩散模型的强大潜力，从此打败此前在图像生成领域统治多年的GAN模型；而后受益于Stable Diffusion开源，扩散模型逐步成为图像生成领域的主流模型。



资料来源：西南证券

2020-2023年图像生成领域技术进展

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
2020					DETR	DDPM				DDIM; Vison Transformer		
2021	CLIP; DALL-E		Swin Transformer		CogView						MAE; Swin Transformer v2; NUWA	Latent- Diff ; GLIDE
2022	BLIP	MaskGIT	Make-A-Scene	DALL-E2 ; CogView2	Imagen	Parti	NUWA-Infinity	Stable Diffusion; BEiT-3; Midjourney V3			eDiff-I ; Midjourney V4	
2023	BLIP2; MUSE		VisualChatGPT; GPT4; Midjourney V5; GigaGAN	SAM			SDXL		DALL-E3		SDXL Turbo	

资料来源：CSDN，西南证券整理

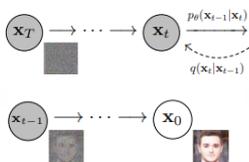
1.2.2 扩散模型：发展历程

- 扩散模型持续迭代，助力应用崛起。2020年，DDPM的提出标志着基础扩散模型的出现；21年5月OpenAI发表论文《Diffusion Models Beat GANs on Image Synthesis》，标志扩散模型打败GAN；2023年，以Midjourney、Adobe Firefly等应用涌现，图像生成领域热度高涨。

扩散模型在图像生成领域中的发展历程

1) 模型提出

- 核心事件：DDPM模型于2020年6月推出，是图像生成领域最早出现的扩散模型。DDPM首次将去噪扩散概率模型应用到图像生成任务中，奠定了扩散模型在图像生成领域应用的基础。
- 论文：Denoising Diffusion Probabilistic Models



2) 模型优化

- 核心事件：图像生成的速度和质量由扩散模型中的采样器控制和决定。目前，Euler、SDE、DPM-Solver++和Karras等加速采样方法的出现有助于大力提升模型在保证生成质量的前提下快速采样能力。
- 论文：Score-Based Generative Modeling through Stochastic Differential Equations



仅20步采样

3) 超越GAN

- 核心事件：2021年5月OpenAI提出在扩散过程中使用显式分类器进行引导，展示扩散模型的强大潜力，打败图像生成领域统治多年的GAN模型，扩散模型逐步成为图像生成领域的主流。
- 论文：Diffusion Models Beat GANs on Image Synthesis

① BigGAN-deep模型结果



② OpenAI扩散模型结果



4) 爆发

- 核心事件：21年1月CLIP预训练模型发布，该技术而后与扩散模型持续结合，推动扩散模型21-22年在文生图领域爆发。如OpenAI GLIDE、DALL-E、Google Imagen、Stable Diffusion。
- 论文：Learning Transferable Visual Models From Natural Language Supervision



5) 模型再优化

- 核心事件：自22H2以来，DreamBooth(22.08)、LoRA(23.01)和ControlNet(23.02)等扩散模型的再学习方法持续推出，针对不同任务对扩散模型进行优化。
- 论文：DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation



6) 应用崛起

- 核心事件：众多图像生成应用应运而生。Midjourney V5；Stability AI公司开发的Dream Studio图像生成工具箱；Adobe Firefly；百度推出文心一格 AI创作平台；阿里巴巴达摩院推出通义文生图大模型等。



资料来源：《扩散模型从原理到实践》，西南证券整理

1.2.2 扩散模型：基本原理

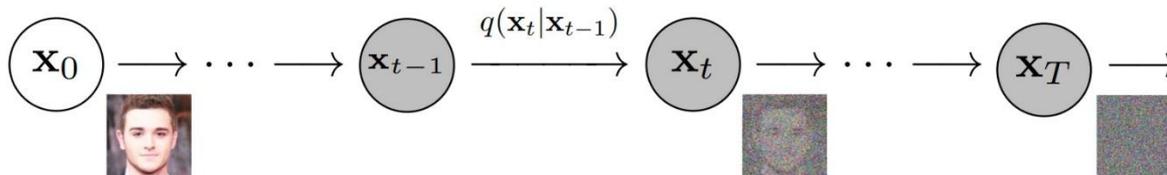
- ❑ **扩散模型 (Diffusion Models) 基本原理**：扩散模型包括前向和反向两个扩散过程，扩散过程是指的对数据逐渐增加高斯噪音直至数据变成随机噪音的过程。
- **前向扩散过程 (Forward Diffusion Process)**：指给图片不断添加噪声的过程，前向过程逐步向真实图片添加噪声最终得到一个纯噪声。对于训练集中的每张图片，都能生成一系列的噪声程度不同的加噪图片；在训练时，不同程度的噪声图片+生成它们所用的噪声是实际的训练样本。
- **反向扩散过程 (Reverse Diffusion Process)**：指给图片去噪的过程，逐渐还原真实数据或样本。

扩散模型的前向过程与反向过程

➢ **前向扩散过程 (Forward Diffusion Process)** → 给图片添加高斯噪声

➢ 前向过程在原始输入图像上逐步添加随机噪声，该噪声服从高斯分布，直至得到趋于纯粹噪声的图像。

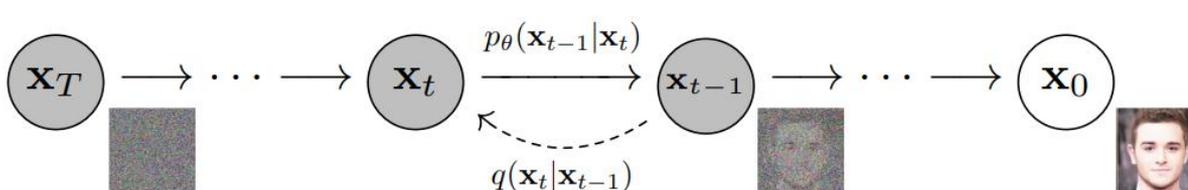
$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1})$$



➢ **反向扩散过程 (Reverse Diffusion Process)** → 给图片去除高斯噪声

➢ 反向过程是不断去除噪声的过程，给定一个噪声图片，对它一步步的去噪还原，直至将原始图像给还原出来。

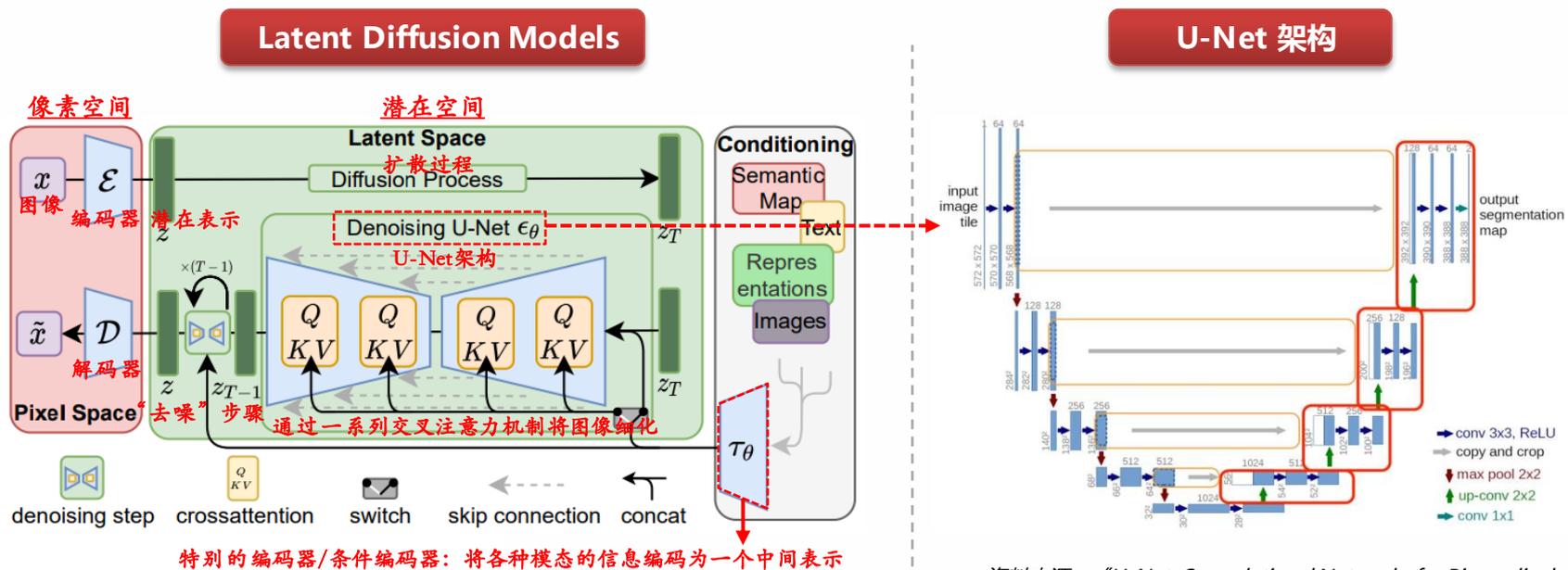
$$p_{\theta}(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t)$$



资料来源：《Denoising Diffusion Probabilistic Models》，西南证券整理

1.2.2 扩散模型：两次关键迭代

- ❑ **扩散模型第一次关键迭代**：在像素空间(Pixel Space)之外引入潜在空间(Latent Space)。扩散模型最初是基于像素空间的Pixel Diffusion，需要处理高分辨率图像等高维 (high dimensional) 数据，导致模型的处理速度较慢。2021年12月，论文《High-Resolution Image Synthesis with Latent Diffusion Models》发表，在扩散模型中引入潜在空间，标志着扩散模型第一次关键迭代。
- ❑ **Latent Diffusion相较于Pixel Diffusion的改进**：Latent Diffusion的基本思想是首先将高维数据降维至feature，再在feature上进行diffusion，完成后将特征空间最后投射至图像空间。因此，Latent Diffusion通过将数据处理成低维 (low dimensional) 数据 (图像的latent space比pixel space更小)，且更加关注数据的重要语义位，能够更好地在“降低复杂度”和“保留细节度”之间进行平衡，最终使模型的计算效率更高。



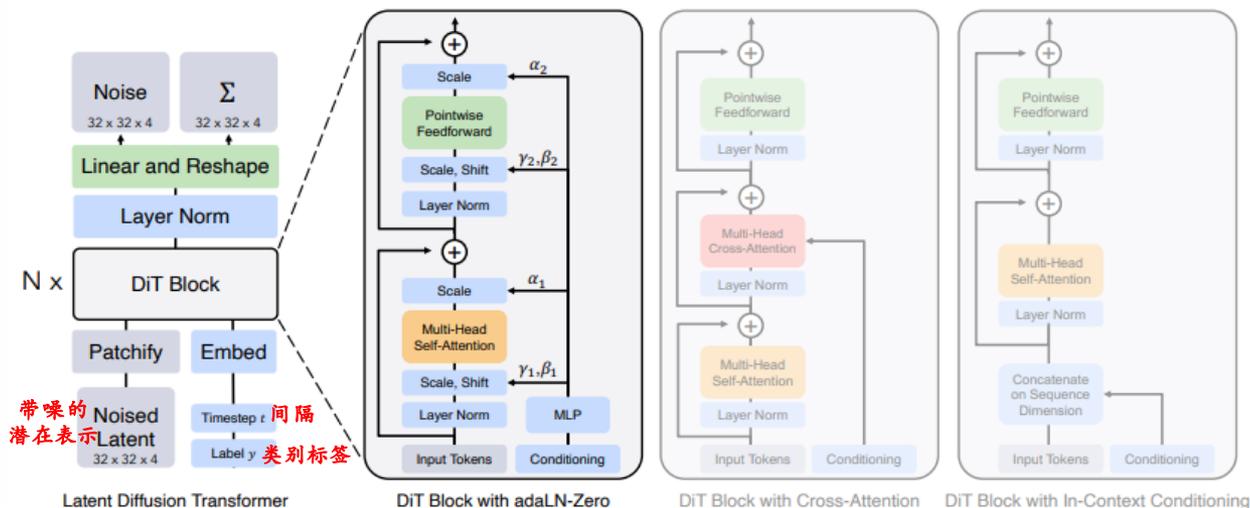
资料来源：《High-Resolution Image Synthesis with Latent Diffusion Models》，西南证券整理

资料来源：《U-Net: Convolutional Networks for Biomedical Image Segmentation》，西南证券整理

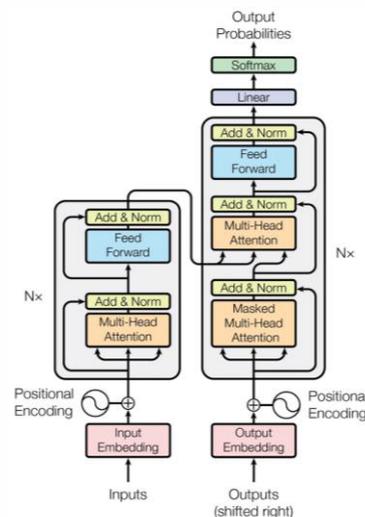
1.2.2 扩散模型：两次关键迭代

- ❑ **扩散模型第二次关键迭代**：在Latent Diffusion基础上将U-Net架构转换成Transformer。扩散模型中的经典架构是U-net网络，大多数扩散模型会基于U-Net结构进行一些变体，U-Net架构主要用于接收“带噪”的输入并预测噪声，以实现“去噪”。2023年3月，Meta发布论文《**Scalable Diffusion Models with Transformers**》，简称DiT (Diffusion Transformer)，将U-Net架构转换成Transformer架构，代表扩散模型的第二次迭代。
- ❑ **Transformer架构相较于U-Net架构的优点**：Transformer架构scalability能力更强、更适合做生成。未来，在扩散模型的技术路线中，更多地结合Transformer架构将成为一大重要趋势。

The Diffusion Transformer (DiT) 架构



Transformer 架构



资料来源：《Scalable Diffusion Models with Transformers》，西南证券整理

资料来源：《Attention Is All You Need》，西南证券整理

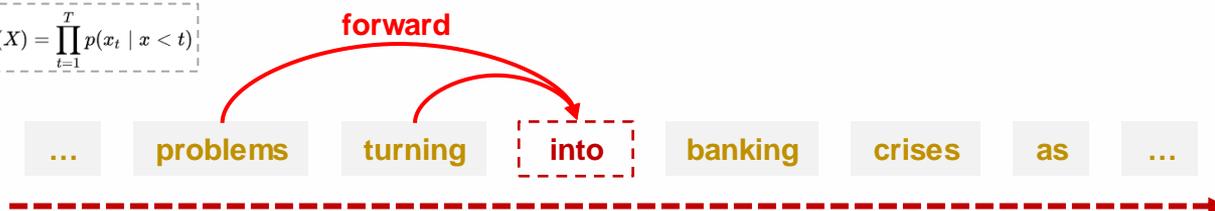
1.2.3 自回归模型：基本原理

- ❑ **自回归模型 (Autoregressive Models) 基本原理**：在数学模型和算法中，自回归模型的核心思想是**基于同一变量的历史值预测未来值**。自回归作为一种时间序列模型，通过使用以前时间步长 (previous time steps) 的观测值作为回归方程的输入，来预测下一个时间步长 (next time step) 的值。
- ❑ **自回归模型具备单向性特征，即前向性或后向性**。在自回归语言模型中，通过利用上 (下) 文词，估计文本语料库的概率分布，预测下 (上) 一个词，**适用于自然语言的生成式任务**，如摘要、翻译和抽象问答，并在**长文本生成**方面取得较大成功。

自回归语言模型的前向与反向过程

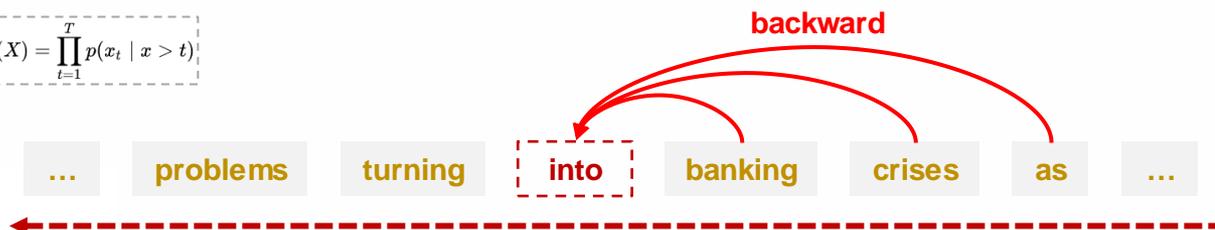
➤ **前向过程**：自回归模型利用上文词，估计文本语料库的概率分布，预测下一个词

$$\text{已知 } x_t \text{ 之前的序列, 预测 } x_t \text{ 的值: } p(X) = \prod_{t=1}^T p(x_t | x < t)$$



➤ **后向过程**：自回归模型利用下文词，估计文本语料库的概率分布，预测上一个词

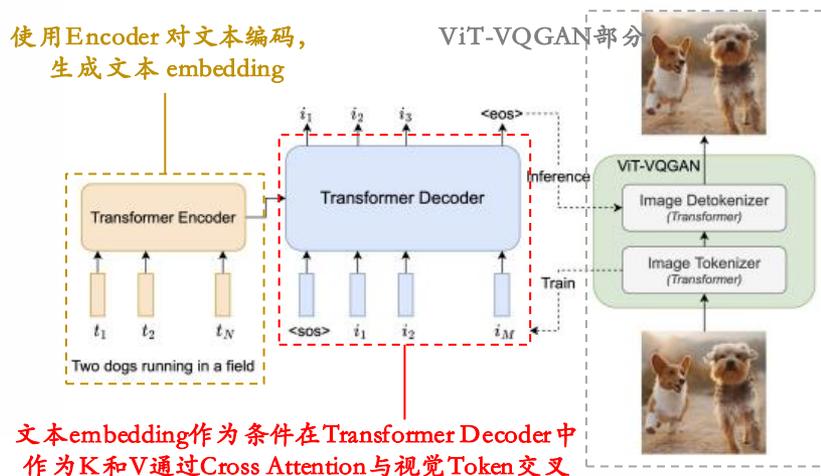
$$\text{已知 } x_t \text{ 之后的序列, 预测 } x_t \text{ 的值: } p(X) = \prod_{t=1}^T p(x_t | x > t)$$



1.2.3 自回归模型：代表模型

- ❑ **自回归文生图代表模型——Parti**：2022年6月22日，谷歌发表论文《Scaling Autoregressive Models for Content-Rich Text-to-Image Generation》，并推出Parti模型（Pathways Autoregressive Text-to-Image model）。该模型受LLM启发，将文本到图片的生成视为序列到序列的建模问题，使用Encoder+Decoder的Transformer架构对文本进行编码以及生成图像Token序列，并将参数量扩展到20B，以获得更好的图像-文本一致性、提升图像生成质量。
- ❑ **自回归文生视频代表模型——VideoPoet**：2023年12月21日，谷歌发表论文《VideoPoet: A Large Language Model for Zero-Shot Video Generation》，并推出VideoPoet模型。该模型强调LLM在视频生成领域依然可以发挥重要作用，能够在动作生成上能够保持较好的一致性、在大范围移动状态下能够保持较好的连贯性，并实现多任务的一站式处理。

谷歌文生图自回归模型—Parti



资料来源：《Scaling Autoregressive Models for Content-Rich Text-to-Image Generation》，西南证券整理

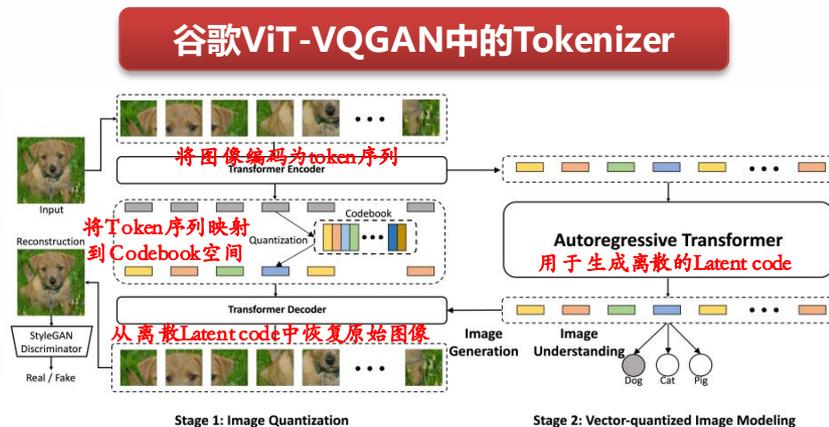
谷歌文生视频自回归模型—VideoPoet



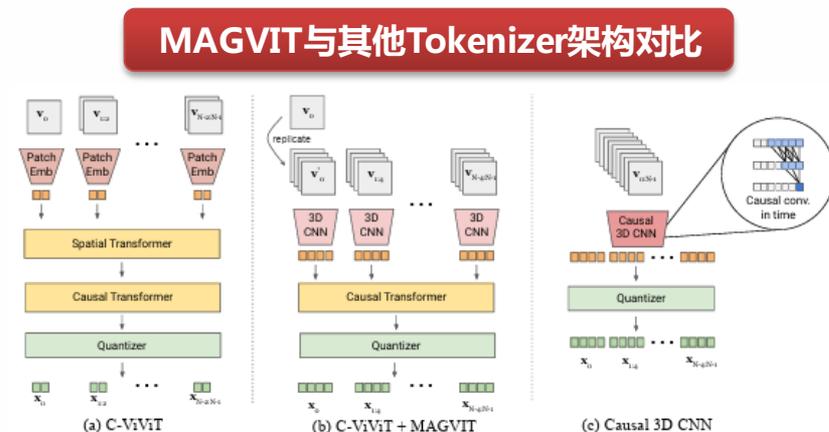
资料来源：《VideoPoet: A large language model for zero-shot video generation》，西南证券整理

1.2.3 自回归模型：Tokenizer的重要性

- ❑ **Tokenizer将文本转化为tokens并建立互联，帮助模型对视觉生成任务的理解。** Tokenizer是指将文本或序列转化为标记 (tokens) 的工具或算法。在自然语言领域，Tokenizer通常将文本转换成tokens；而在视觉领域，visual tokenizer通常将图像转化为patches。在生成模型中，Tokenizer的作用是将连续的输入序列转换为离散的tokens，并建立tokens之间的相关性，使tokens可以更好地被模型理解和处理，帮助模型更准确地完成任务，因此Tokenizer是视觉生成模型中的重要环节。
- ❑ **Image Tokenizer—ViT-VQGAN**：2022年6月5日，谷歌发表论文《Vector-quantized Image Modeling with Improved VQGAN》并提出分词器ViT-VQGAN，而后将其用于Parti模型中，将图片编码为离散token序列，再用其重建图片token序列已生成高质量图像。
- ❑ **Video Tokenizer—MAGViT-v2**：2023年10月9日，谷歌发表论文《Language Model Beats Diffusion: Tokenizer is key to visual generation》，并提出分词器MAGViT-v2 (Masked Generative Video Transformer)，**强调Tokenizer对视觉生成的重要性，指出一个更好的Tokenizer在接入语言模型后有望获得比扩散模型更好的效果。**



资料来源：《Vector-quantized Image Modeling with Improved VQGAN》，西南证券整理



资料来源：《Language Model Beats Diffusion: Tokenizer is key to visual generation》，西南证券整理

1.3 未来趋势：transformer优势凸显，模型构建有望呈现多种方式

□ **Transformer扩展能力突出，扩散模型与其持续结合。**当前，为构建性能更优、效率更高的图像和视频生成模型，已出现多种结合Diffusion Model和Transformer架构的文生视频模型。例如，谷歌于2023年12月发布与李飞飞老师和其学生合作的文生视频模型——W.A.L.T、OpenAI于2024年2月发布的文生视频模型——Sora。扩散模型通过引入Transformer架构，发挥其扩展（scaling）能力，且transformer已在语言建模、计算机视觉和图像生成等多个领域展示其显著的扩展能力，未来同样有望在视频生成上发挥更大作用。随着技术研究的不断深入，各家可能会**利用每个架构最擅长的部分，呈现多种方式构建模型**，扩散模型和transformer架构有望持续结合。

Sora基于Diffusion+transformer



Sora是基于Transformers+扩散模型

Scaling transformers for video generation

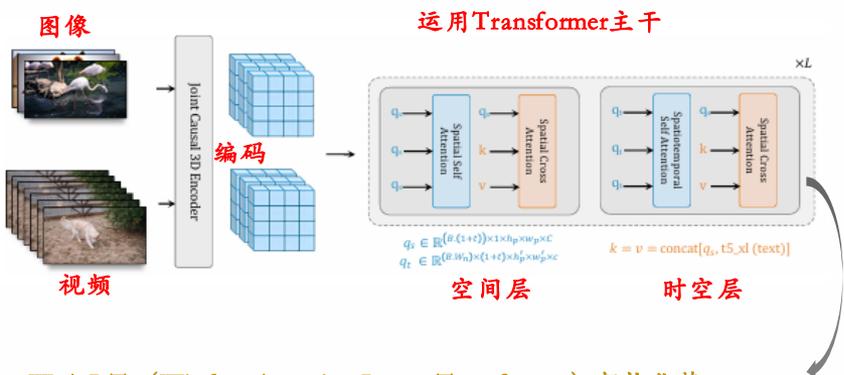
Sora is a diffusion model^{21,22,23,24,25}, given input noisy patches (and conditioning information like text prompts), it's trained to predict the original "clean" patches. Importantly, Sora is a diffusion transformer²⁶. Transformers have demonstrated remarkable scaling properties across a variety of domains, including language modeling,^{15,16} computer vision,^{15,16,17,18} and image generation.^{27,28,29}



Sora技术报告指出：Transformers已经在语言建模、计算机视觉和图像生成等多个领域展示显著的扩展能力

资料来源：OpenAI官网，西南证券整理

W.A.L.T基于Diffusion+transformer



W.A.L.T (Window Attention Latent Transformer) 架构优势：

- ① local window attention有助于显著降低计算需求；
- ② 有利于联合训练，空间层处理图像和视频帧，时空层用于建模视频中的时间关系。

资料来源：《Photorealistic Video Generation with Diffusion Models》，西南证券整理

目 录

1 从Sora看文生视频模型的技术路径

- 1.1 发展现状：视觉生成热潮兴起，路径探索仍在继续
- 1.2 路径对比：扩散模型占据主流，自回归模型潜力可期
- 1.3 未来趋势：transformer优势凸显，多种模型构建方式有望呈现

2 从Sora看文生视频模型的最新能力

- 2.1 模型背景：团队成员年轻有为，相关技术积累深厚
- 2.2 理解能力：领会并细化提示词，学习及模拟物理世界
- 2.3 生成能力：视频长度优势明显，连贯性及可控性提升
- 2.4 其他能力：图生视频，视频扩展/编辑/拼接/模拟等

3 从Sora看文生视频模型的影响

- 3.1 算力端：参数扩展提升训练算力消耗，应用落地拉动推理算力增长
- 3.2 应用端：文生视频模型拓宽应用广度，前期有望以助手角色落地
- 3.3 其他：云服务将成算力重要补充，多媒体传输带动光通信需求

2.1.1 团队背景：研发实力突出，优秀人才聚集

- **研发团队年轻，优秀人才聚集。**根据OpenAI Sora技术报告，本篇报告作者共13位。其中，Aditya Ramesh、Tim Brooks和Bill Peebles是团队的核心成员，Aditya作为图像生成模型DALL-E的提出者，主导2021年至2023年DALL-E模型的三次迭代，相关领域研究深入，且团队成员优秀，包括众多年轻人才，较高的人才密度是打造优质模型的基础。

OpenAI Sora团队核心成员履历

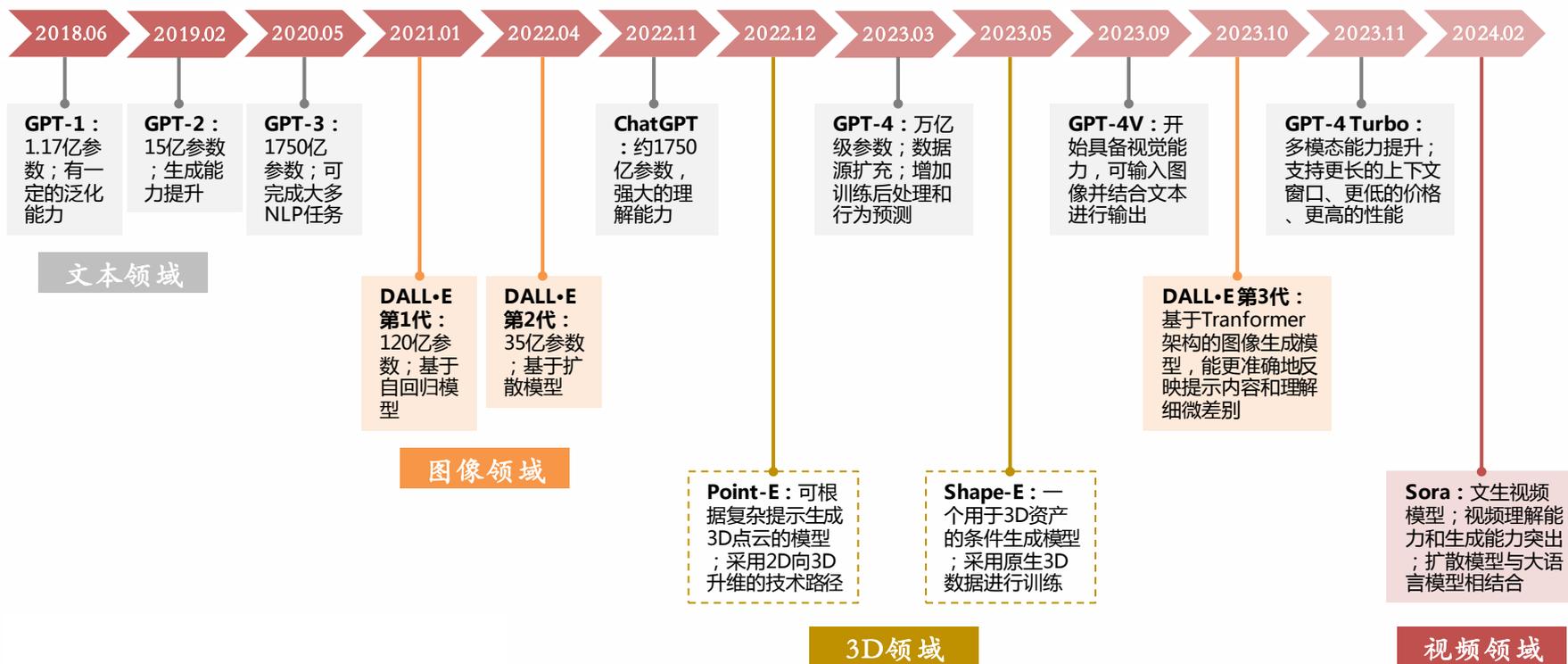
核心作者	主要研究成果	个人介绍	推特账号
Aditya Ramesh	<p>【语言领域】 2020年5月28日 《Language models are few-shot learners》</p> <p>【图像领域】 2021年1月5日 《DALL-E: Creating images from text》</p> <p>【图像领域】 2022年4月13日 《Hierarchical text-conditional image generation with CLIP latents》</p> <p>【图像领域】 2023年10月20日 《Improving Image Generation with Better Captions》</p> <p>【视频领域】 2024年2月15日 《Video generation models as world simulators》</p>	<p>DALL·E 第1代至第3代</p> <p>Sora团队负责人；本科就读于纽约大学；主导三代DALL-E的研究</p>	<p>https://twitter.com/model_mechanic</p>
Tim Brooks	<p>【视频领域】 2022年6月9日 《Generating Long Videos of Dynamic Scenes》</p> <p>【图像领域】 2023年1月18日 《InstructPix2Pix: Learning to Follow Image Editing Instructions》</p> <p>【图像领域】 2023年10月20日 《Improving Image Generation with Better Captions》</p>	<p>Sora团队核心成员&主要作者；本科就读于卡内基梅隆大学，主修逻辑与计算；博士就读于加州大学伯克利分校AI实验室；曾在Facebook软件工程部门、Google Pixel手机部门、英伟达实习或工作过</p>	<p>https://twitter.com/tim_brooks</p>
Bill Peebles	<p>【图像领域】 2022年4月5日 《GAN-Supervised Dense Visual Alignment》</p> <p>【图像领域】 2023年3月2日 《Scalable Diffusion Models with Transformers》</p>	<p>Sora团队核心成员&主要作者；本科就读于麻省理工学院；博士就读于加州大学伯克利分校AI实验室；曾在Facebook人工智能研究中心FAIR、Adobe Research、英伟达实习过</p>	<p>https://twitter.com/billpeeb</p>

资料来源：OpenAI官网，西南证券整理

2.1.2 时间进程：AGI技术积淀已久，视觉生成领域加速推进

- 前期工作积累深厚，研发进程加速推进。Sora模型的诞生离不开GPT和DALL·E系列模型的积淀，GPT语言模型从2018年开始迭代，第一代DALL·E文生图模型于2021年1月正式发布论文，优质基座模型和文本-图像配对的优质数据集奠定前期基础，Sora模型最终于2024年2月16日对外宣布。

OpenAI在各模态领域的模型布局

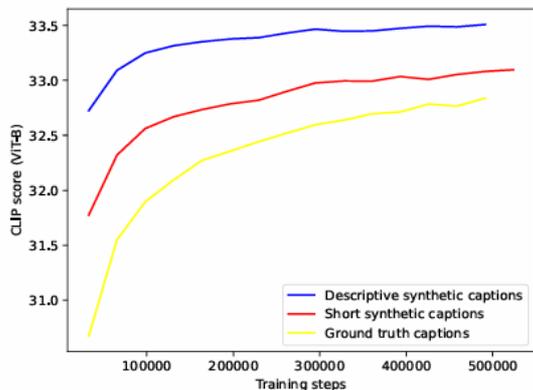


2.2.1 强大的理解能力：领会并细化提示词，学习及模拟物理世界

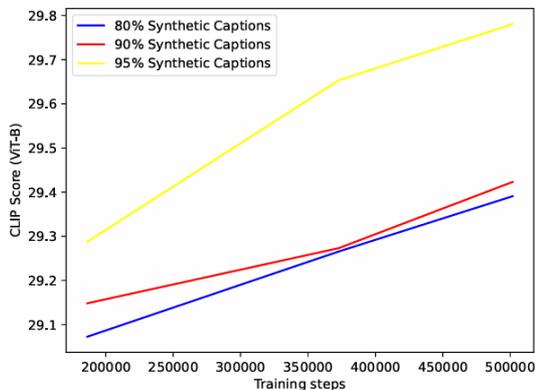
❑ 为什么Sora可以较好地理解Prompt提示词？

- **优秀的基座模型：DALL·E帮助文本-视频进行高质量配对+GPT用于Prompt的细化处理。** GPT、DALL·E模型迭代至今，已呈现对语言、视觉数据和世界较强的理解能力。
- ① **DALL·E帮助文本和视频的高质量配对：** OpenAI技术报告指出，研发团队将DALL·E 3的文字描述技术应用到视频训练上，训练出一个较强的文字描述模型，并在训练Sora时通过使用大量的与视频相对应的高准确度描述性文字，提高文生视频的质量。
- ② **GPT用于改写和细化Prompt：** 团队利用GPT系列模型将简短的Prompt转换成详细的文字描述，将Prompt详细化、细节化，用于模型训练中，使Sora能够更好地遵循用户提示并准确地生成视频。

详细的文字描述有助于高质量图像生成



在“更详细的文本和图像配对”上训练的图生图模型的CLIP分数更高



详细的描述性合成caption比例越高，训练的图生图模型的CLIP分数越高

GPT模型将Prompt提示词详细化



A bird scaring a scarecrow.

GPT可帮助补充缺失的细节、消除复杂关系的歧义，图像质量提升

A large, vibrant bird with an impressive wingspan swoops down from the sky, letting out a piercing call as it approaches a weathered scarecrow in a sunlit field. The scarecrow, dressed in tattered clothing and a straw hat, appears to tremble, almost as if it's coming to life in fear of the approaching bird.

2.2.1 强大的理解能力：领会并细化提示词，学习及模拟物理世界

- 争议：Sora是否能够理解事物在物理世界中的存在方式？
 - Sora在一定程度上展示出其可能具备理解物理现象和物理规律的能力。根据OpenAI官网的展示视频，两艘海盗船在咖啡的波浪中起伏，咖啡的流动、船只周围的泡沫都十分自然，可以看出，尽管该场景实际上不存在于现实世界，但Sora依然可以根据流体力学等物理规则实现Prompt的要求。
 - Nvidia高级研究科学家Jim Fan观点：Sora很可能是一个基于数据驱动的物理引擎，是对许多世界的真实或幻想的模拟，并且模拟渲染直观的物理、推理和基础。Sora必须学习一些隐式形式的文本到3D的转换、光线追踪渲染和物理规则，以便尽可能准确地对视频像素进行建模。
 - Transformer架构下的Scaling law对理解物理现象和规则有望持续有效。当模型变大、数据量更多、视觉场景的分辨率越来越高，Sora所理解的规则的颗粒度将会越来越细，对物理现象和规律的把控越来越精准，并支持更长上下文的学习，使视频在更长的时间维度下做到一致与合理。

Sora模型生成视频展示1



- 一只猫叫醒熟睡的主人要求吃早餐，猫咪出现猫捏/踩奶的动作。

资料来源：OpenAI社交媒体账号，西南证券整理

Sora模型生成视频展示2



- 两艘海盗船在咖啡的波浪中起伏，咖啡的流动、船只周围的泡沫十分自然。

资料来源：OpenAI社交媒体账号，西南证券整理

2.2.1 强大的理解能力：领会并细化提示词，学习及模拟物理世界

- 争议：Sora是否能够理解事物在物理世界中的存在方式？
 - Sora在理解物理世界方面仍存在缺陷。OpenAI Sora技术报告指出，Sora当前仍存在一些局限，如，不能准确地模拟许多基本交互的物理现象，尚不能对所有情况进行完全准确的建模。
 - Transformer架构和扩散模型在学习和模拟物理世界的过程中存在一些技术限制：1) 相关性vs因果律：transformer可以建立tokens间的强相关性，但相关性不等于因果关系，Sora模型生成的视频可以在帧与帧之间做到较好的连贯性，但不意味着符合因果逻辑，所以在“奶奶吹蜡烛”的展示视频中，画面十分连贯，却在吹完蜡烛后火苗并没有熄灭。2) 稳恒态vs临界态：Sora所用的扩散模型适合学习和生成稳定状态下的物体，但是较难理解和识别物体发生显著变化时的临界状态，所以在“玻璃杯摔倒”的展示视频中，杯子摔倒没有出现骤然破碎的状态。

Sora模型生成视频展示3



- 一位奶奶吹生日蛋糕上的蜡烛后，蜡烛并没有被熄灭。

资料来源：OpenAI社交媒体账号，西南证券整理

Sora模型生成视频展示4



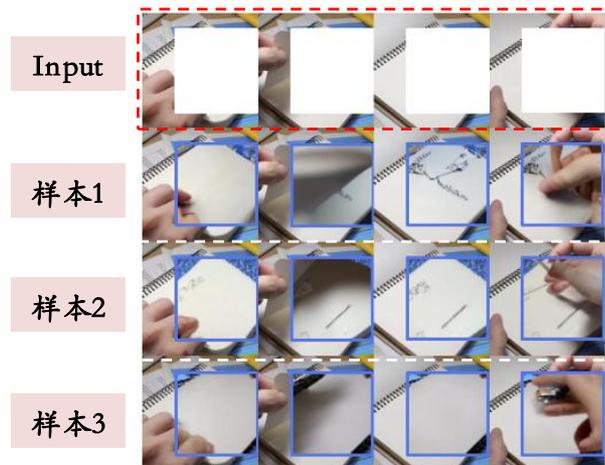
- 玻璃杯倒下后并没有破碎，而是部分杯体变成液体状态。

资料来源：OpenAI官网，西南证券整理

2.2.1 强大的理解能力：领会并细化提示词，学习及模拟物理世界

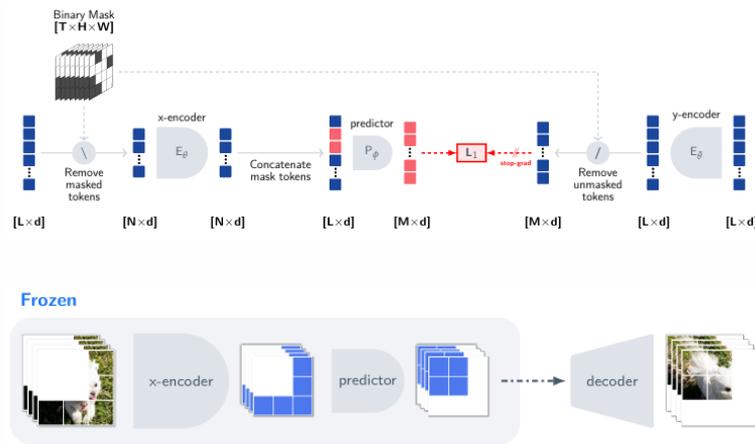
- 争议：Sora是否能够理解事物在物理世界中的存在方式？
- **Meta V-JEPA模型**：2024年2月14日，Meta发表论文《Revisiting Feature Prediction for Learning Visual Representations from Video》并推出V-JEPA模型（Video Joint Embedding Predictive Architectures）。不同于视频生成模型Sora，V-JEPA模型通过学习图像和视频的表示，主要用于**预测视频缺失的部分或者被遮住的部分**，目标是希望**从内在学习并理解物理世界的概念**。
- **Yann LeCun观点**：大多数根据提示生成的逼真的视频并不表明模型能够理解物理世界。生成模型与**世界模型的因果预测**非常不同。对于生成模型来说，可信视频(plausible videos)的空间非常大，因此只需生成一个样本即可算作成功；而一个真实视频的合理延续(plausible continuations of a real video)的空间要小得多，生成一个有代表性的块是一个更艰难的任务，特别是有条件限制时。

Meta V-JEPA模型设计架构



第一行：被遮挡的视频用作V-JEPA模型的输入

第二行至第四行：蓝色框线内包含来自覆盖在原始视频上的解码器decoder出来的各种样本。模型生成的样本与输入不一定完全匹配。V-JEPA不仅预测空间上的信息，同时能够捕捉随时间变化的一致运动



资料来源：《Revisiting Feature Prediction for Learning Visual Representations from Video》，西南证券整理

2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

□ **长度**：Sora可以生成长达约60s的视频。

- **Sora**：根据OpenAI官网展示的Sora模型生成的视频，假设展示视频可以代表Sora的一般性表现，最短的视频为8s，最长的视频为59s，多数视频的长度集中在20s左右的水平。
- **Pika 1.0**：2023年11月29日发布，生成的视频一般为3s，多的为7s。
- **Runway Gen-2**：生成的视频长度在4s~18s之间不等。Gen-2最初于2023年3月发布，可生成4s时长的视频；23年8月，公司在Gen-2中增加新选项，可视频延长至18s。

当前主要文生视频产品对比

产品	公司	技术路径	时长	画质	发展阶段
Sora	OpenAI	基于Transformer的扩散模型	8s~60s	视频分辨率可达1920x1080p； 图片分辨率可达2048x2048p	2024年2月16日发布，目前处于测试阶段，仅面向“红队”成员和少数创作者开放。
Pika 1.0	Pika labs	扩散模型	3s / 7s	最高可达2K画质	2023年11月29日发布，12月开放测试资格，兑现了文生图、文生视频、图生视频、视频扩展、区域修改等能力。24年1月，上线付费计划，分为标准和专业两个档次，分别是10美元和60美元。
Runway Gen-2	Runway	扩散模型	4s~18s	由静态图像生成的Gen-2视频的分辨率可达2816×1536p	2023年3月发布，对Gen-1进行改进，允许用户输入文本提示，生成4s时长的视频，或上传图片让其添加动作。23年8月，视频时长可延长至18s。23年9月，增加“导演模式”新功能，允许用户在生成的视频中选择“摄像机”运动的方向和强度/速度。
Stable video diffusion	Stability.AI	扩散模型	提供14或25帧，可在3帧/s~30帧/s之间调整帧率	576x1024p	2023年11月22日发布，最新版本为SVD、SVD-XT，目前仅限于研究，不用于实际商用；目前产品在GitHub中开源，Discord社区拥有约32W位成员，暂未开放收赛渠道，Stable Video网页端已启动内测。

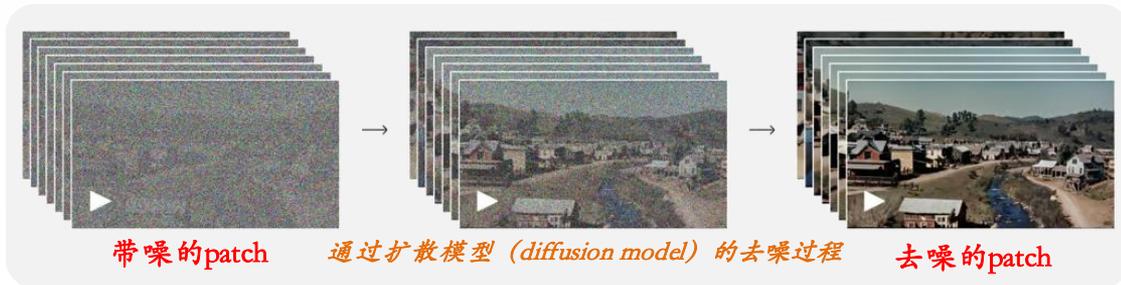
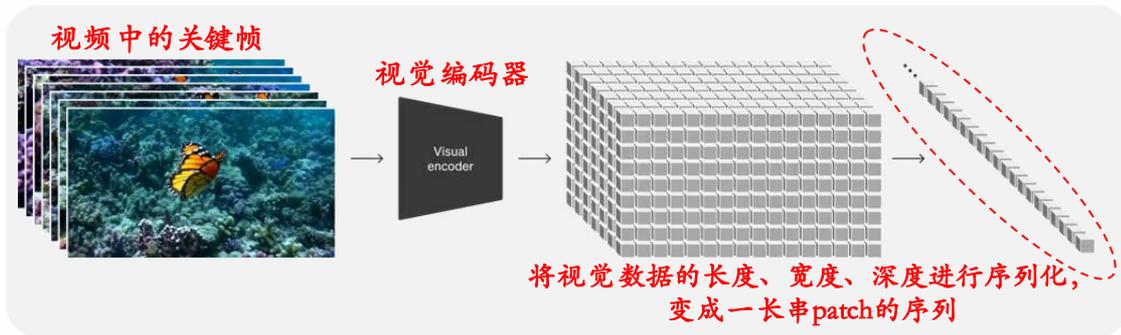
资料来源：各公司官网，西南证券整理

2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

❑ 为什么Sora能够生成的长达60s的视频？

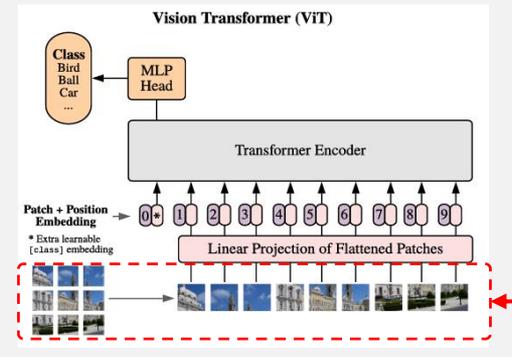
→ Patch的表达方式（前提）+ transformer时空信息的处理能力（带来时序概念）+ transformer的scaling扩展能力（能力）。Sora将视觉数据统一转化为patch的表示方法，在transformer架构下，patch和patch之间的相关性既包含空间上的相关性，也包含时间上的相关性，而transformer可以更好地处理视频中各个关键帧像素信息的相关性。因此，在更多的patch下，Sora模型可以把一个更长的视频完整地表达出来，而transformer的scale up的能力也更加make sense。

Sora模型将视觉数据转化为Patch的表达方式



❑ 什么是Patch？

- 将图片分割后的小块。为方便理解，下图展示的是显式的patch，与之对应的是隐式（latent）的patch，Sora其实也正是用的latent patches。
- Patches的思想最一开始是源于Google深度学习团队的ViT-Vision transformer算法。



资料来源：OpenAI官网，《An image is worth 16x16 words: Transformers for image recognition at scale》西南证券整理

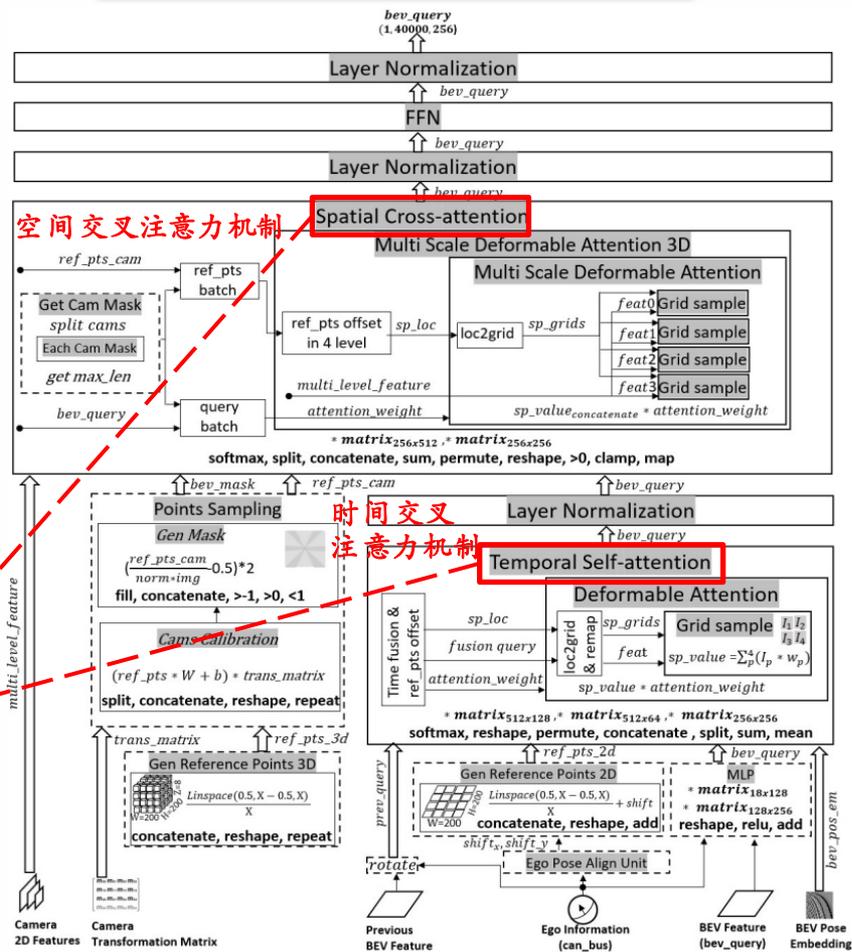
2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

Transformer对时空信息的处理能力：

- ① Transformer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。
- ② Transformer在处理大规模数据量场景上具备优势，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。
- ③ Transformer架构引入**时间和空间注意力机制**，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**。

- 以自动驾驶中的BEVformer编码器为例，Transformer具备时空注意机制。
- 相较于纯扩散模型而言，Transformer架构在时空信息编码和解码的能力上十分突出，在文生视频领域应用Transformer架构，可以更好地处理视频中各个关键帧像素信息的相关性，例如画面中空间语义的相关性、序列中时间语义的相关性。**强大的时空信息相关性的处理能力有助于模型可以生成更长时间维度的视频。**

自动驾驶领域中BEVformer的时序概念



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

2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

□ Transformer的scaling扩展能力：

- **Scalable/Scale up能力**：即“大力出奇迹”，模型越大、训练数据越多，效果越强，出现智能涌现的确定性越高。从模型参数量级来看，截至目前，扩散模型的参数规模普遍在billion（十亿）级别，而自回归模型的参数通常是扩散模型参数量的10倍以上。
- 市场大多认为扩散模型是图像和视频生成领域的技术主流，却容易忽视语言模型中transformer架构scale up的能力。OpenAI技术报告指出，Sora是基于扩散模型，但更强调，**Sora是一个基于Transformer架构的扩散模型，其优秀的生成能力离不开Transformer架构优秀的scaling特性。**

Sora将Scaling transformers运用到视频生成中的效果

- 以Sora技术报告中“带着蓝色帽子的小狗”为例



Base compute



4x compute



32x compute

- 可以看出，在计算扩展(scaling)的情况下，视频效果显著提升

◆ Base compute

◆ 如何理解 4x compute ?

- 例如，在base compute的基础上，将模型参数量扩大2倍、训练数据集扩大2倍

◆ 如何理解 32x compute ?

- 例如，在4x compute的基础上，再将模型参数量扩大4倍、训练数据集扩大2倍

2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

- ❑ **复杂度**：能够生成包含多个角色、多种主题、多类运动形态的复杂场景。
- ❑ **逼真度**：能够呈现更多的视觉细节，具备更高清的画质，为用户提供更好的视觉体验。生成的视频的画质分辨率可以达到1920x1080p。生成的图片的画质分辨率可以达到2048x2048p。
- ❑ **连贯性&一致性**：在角度移动和切换情况下，人物和场景等元素在3D空间中的位置关系能够保持一致的运动；可以生成同一角色的多个镜头，保持其在整个视频中的外观，当物体短暂地被遮挡或离开框架后仍然能够恢复与先前一致的物体特征。
- ❑ **可控性**：在某一Prompt基础上只改动一个关键词仍能生成优质的视频，具备较好的可控性。

Sora模型连贯性&一致性示例



① 小狗原先的特征 → ② 小狗即将被遮挡



④ 小狗再现后特征仍保持一致 ← ③ 小狗被遮挡

Sora模型可控性示例

a woman wearing blue jeans and a white t-shirt taking a pleasant stroll in Mumbai, India during a beautiful sunset

在Prompt中只改动少量关键词，仍可快速生成优质视频



a woman wearing blue jeans and a white t-shirt taking a pleasant stroll in Mumbai, India during a colorful festival



资料来源：OpenAI官网，西南证券整理

2.2.2 优秀的生成能力：视频长度优势明显，连贯性及可控性提升

- 为什么Sora能够在视频生成的逼真度/连贯性/一致性方面能够取得较大进展？
 - 建立优质的视觉数据集和文本-视频配对集：得益于前期DALL·E 3的相关工作，Sora拥有具备高关联性、高质量的文本-图像-视频配对集，使不同视频帧之间能够保持较高的一致性。
 - 对视觉训练数据的处理能力：通过降维的方法保留原始视觉数据格式。OpenAI通过降维的方法，保留原始图像的纵横比和分辨率，使画面完整度更高，这一改进有助于模型更准确地捕捉视觉数据的真实本质，使Sora能够在不用调整大小或填充画面的情况下，高效地处理各种视觉数据。此前的视频生成模型的工作（如GAN、Autoregressive Transformers、Diffusion Models）通常局限于视觉数据的尺寸大小比如裁剪成标准尺寸、时长为4秒、分辨率为256x256，而Sora可以直接使用数据的原始尺寸进行训练，因此Sora生成的视频可以跨越不同持续时间、高宽比和分辨率，画面构图更加合理，内容逻辑性和连贯性也更好，视频输出质量更高。

Sora模型在视频采样的尺寸方面具备灵活性



- Sora可以采样宽屏1920x1080p、垂直1080x1920p的视频，以及介于两者之间的视频。

Sora模型在画面构图方面更合理



- 左图：在标准尺寸上训练的模型有时会出现物体只有一部分呈现在视频中的问题。
- 右图：Sora基于视频的原始尺寸进行训练，生成的视频较左图有明显改进。

2.2.3 其他能力：图生视频，视频扩展/编辑/拼接/模拟等

- ❑ **图生视频**：可以将图片转化为动图和视频。
- ❑ **视频扩展**：能在时间上向前或向后扩展视频。例如在时间上向后延伸，视频开头不同，但结局相同。
- ❑ **视频编辑**：扩散模型已有众多根据文本提示编辑图像和视频的方法，可改变物体样式和画中环境等。
- ❑ **衔接视频**：使用Sora在两个输入视频之间逐步插值，实现不同的主题和场景视频的无缝过渡。
- ❑ **模拟数字世界**：Sora能够模拟人工过程，例如电子游戏等。

图生视频的能力



静态图片变成动态视频



视频编辑的能力



背景环境从秋季树木枯黄
换成夏季绿树成荫



衔接视频的能力



无人机和蝴蝶实现无缝衔接



资料来源：OpenAI官网，西南证券整理

目 录

1 从Sora看文生视频模型的技术路径

- 1.1 发展现状：视觉生成热潮兴起，路径探索仍在继续
- 1.2 路径对比：扩散模型占据主流，自回归模型潜力可期
- 1.3 未来趋势：transformer优势凸显，多种模型构建方式有望呈现

2 从Sora看文生视频模型的最新能力

- 2.1 模型背景：团队成员年轻有为，相关技术积累深厚
- 2.2 理解能力：领会并细化提示词，学习及模拟物理世界
- 2.3 生成能力：视频长度优势明显，连贯性及可控性提升
- 2.4 其他能力：图生视频，视频扩展/编辑/拼接/模拟等

3 从Sora看文生视频模型的影响

- 3.1 算力端：参数扩展提升训练算力消耗，应用落地拉动推理算力增长
- 3.2 应用端：文生视频模型拓宽应用广度，前期有望以助手角色落地
- 3.3 其他：云服务将成算力重要补充，多媒体传输带动光通信需求

3.1 算力端：参数扩展提升训练算力消耗，应用落地拉动推理算力增长

- ❑ **训练端：模型参数仍有扩展空间，训练算力消耗有望提升。** 从目前市场上的文生视频模型来看，参数大小一般处于1B~20B之间，相较于千亿甚至万亿参数级别的大语言模型而言，模型仍有扩展空间，随着模型持续scaling，训练算力消耗有望持续提升。
- ❑ **推理端：未来应用落地可期，推理算力需求持续增长。** 扩散模型包括大量的去噪步骤，扩散步骤带来的算力消耗相较于大语言所需的推理算力需求更高。此外，在视频生成的过程中，视频需要转换成帧，再转换成patch，因此，随着应用的落地，推理侧所产生的视频帧和patches量将持续增长，从而带动算力需求，且视频越长、清晰度越高、对物理世界的理解越好，算力消耗越大。

文生视频模型参数情况

文生视频模型通常在1B~20B之间

Method	#data	#params.	IS(↑)	FID(↓)	FVD(↓)
CogVideo (En) [23]	5.4M	15.5B	25.27	179.00	701.59
MagicVideo [55]	10M	-	-	145.00	699.00
LVDM [16]	2M	1.2B	-	-	641.80
InternVid [46]	28M	-	21.04	60.25	616.51
Video-LDM [5]	10M	4.2B	33.45	-	550.61
ModelScope [43]	10M	1.7B	-	-	410.00
VideoFactory [44]	-	2.0B	-	-	410.00
Make-A-Video [37]	20M	9.7B	33.00	-	367.23
VidRD [13]	5.3M	-	39.37	-	363.19
Dysen-VDM [7]	10M	-	35.57	-	325.42
PixelDance	10M	1.5B	42.10	49.36	242.82

文生视频模型推理算力的影响因素

推理算力影响因素

视频的Sequence Length较文本模态更长，且视频越长，消耗的算力越大

视频的清晰度越高，算力要求越大

长视频对连贯性和一致性的要求更高

物理规则更好的理解要求更多的推理

Diffusion Model的大量去噪步骤，带来的算力消耗显著高于大语言模型

资料来源：《Make Pixels Dance: High-Dynamic Video Generation》，西南证券整理

3.2 应用端：文生视频模型拓宽应用广度，前期有望以助手角色落地

- **大模型向多模态方向持续发展，AI应用的广度进一步拓宽。**目前，OpenAI已向部分视觉艺术家、设计师和电影制作人提供访问权限，以获取专业的反馈。根据OpenAI内测方向也代表了文生视频模型应用的重点领域，我们认为以Sora为代表的视频生成模型有望给广告业、设计业、电影业、短视频行业、游戏行业带来变化。
- **视频生成模型前期将扮演创作工具的角色，对文娱行业等进行赋能。**当前，以Sora为代表的视频生成模型距离精确控制尚有一定差距，受制于精确控制的能力，Sora未来的落地形式更可能偏向于ChatGPT等文本端工具，以助手的角色赋能文娱等领域，从中短期来看，视频生成模型将更多的作为创作工具对相关行业进行赋能。

文生视频模型对应用领域的影响

细分领域	文生视频潜在赋能方向
短视频	视频生成模型在视频生成时长上较短剧和影视作品来看更能满足实际需求，文生视频模型有望降低短视频创作门槛，提升创作效率，同时为长尾视频内容的生产提供更多的可能性。
广告	主流的广告素材更多地倾向于视频模态，文生视频技术的进步有望提升视频广告生产效率。
IP & 社交	帮助创建虚拟人，创作IP将变得更加简单，成本低优势。
游戏	高性能带宽的提升对游戏的提升效率会更高；可以帮助创建游戏Demo，提升团队讨论效率等。
电影	有望改变电影制作的传统流程，例如减少对物理拍摄、场景搭建、特效制作等资源的依赖，降低电影制作成本，同时使个人创作者和小型团队也有可能制作出具有专业水准的影视作品；AI主要是介入内容生产和编辑的方式，对内容创作流程的改变。
AR/VR	有助于打造虚拟生态，为用户带来沉浸式的体验；但AR/VR场景还需用到空间计算，用户的设备定位对精度要求高。
自动驾驶	自动驾驶的进步仍需视觉系统的进一步发展，视觉系统还需识别地图等三维深度信息，需要更精准地理解和识别世界。
工业	未来有望赋能3D和物理仿真等环节。

资料来源：西南证券

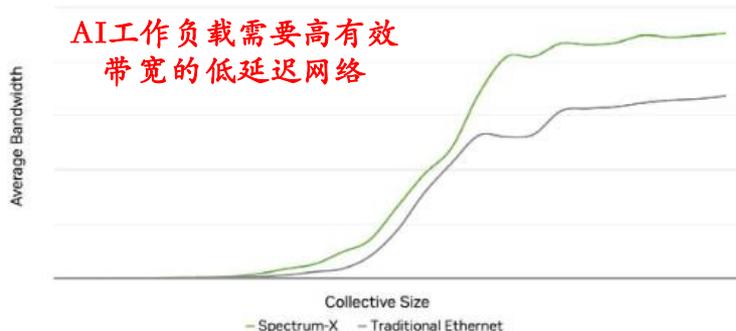
3.3 其他：云服务将成算力重要补充，多媒体传输带动光通信需求

- ❑ **云服务**：模型厂商除购买大量算力以支撑AI服务，同时还需使用云服务作为补充。在推理侧，AI模型的运行大多需要依赖云服务进行生成。未来，随着Sora等更多的文生视频模型开放对外服务，将进一步提高市场对云服务的需求。
- ❑ **光通信**：随着视频生成模型的发展以及未来应用的落地，数据中心之间将进行更多的视频等多媒体信息的传输，对传输带宽的要求和稳定性的要求更高，光通信需求有望大幅提升。为应对未来大模型的升级迭代和大规模应用，需要更高性能的网络带宽和更低的延迟，以支持更复杂的网络需求。目前，交换机的网络容量在数万张卡的级别，为进一步扩大网络容量，支撑更大体量的芯片集群，市场对交换机、光模块、光芯片等基础设施的需求也将持续增长。

英伟达数据中心GPU路线图

Adaptive Routing Performance Chart

All-to-All

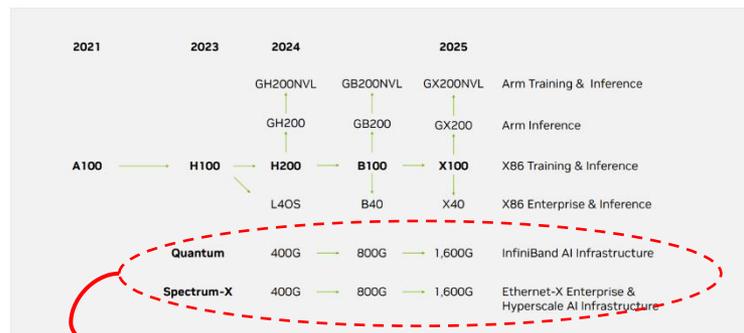


资料来源：英伟达官网，西南证券整理

英伟达数据中心GPU路线图

NVIDIA AI - One Architecture | Train and Deploy Everywhere

From Two-Year Rhythm to One-Year Rhythm | Training & Inference | x86 & Arm | Hyperscale & Enterprise



1.6T光模块需求有望增长

资料来源：英伟达官网，西南证券整理

风险提示

- ❑ 技术进展不及预期风险；
- ❑ 行业竞争加剧风险；
- ❑ 应用开发不及预期风险。



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

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与作者在自身所知知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

《证券期货投资者适当性管理办法》于2017年7月1日起正式实施，本报告仅供本公司签约客户使用，若您并非本公司签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告，本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可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。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，本公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

本报告及附录版权为西南证券所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附录的，本公司将保留向其追究法律责任的权利。



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	田婧雯	高级销售经理	18817337408	tjw@swsc.com.cn
	崔露文	销售副总监	15642960315	clw@swsc.com.cn	张玉梅	销售经理	18957157330	zymyf@swsc.com.cn
	谭世泽	高级销售经理	13122900886	tsz@swsc.com.cn	龙思宇	销售经理	18062608256	lsyu@swsc.com.cn
	汪艺	高级销售经理	13127920536	wyyf@swsc.com.cn	阚钰	销售经理	17275202601	kyu@swsc.com.cn
	李煜	高级销售经理	18801732511	yfliyu@swsc.com.cn	魏晓阳	销售经理	15026480118	wxyang@swsc.com.cn
	卞黎旸	高级销售经理	13262983309	bly@swsc.com.cn				
北京	李杨	销售总监	18601139362	yfly@swsc.com.cn	张鑫	高级销售经理	15981953220	zhxin@swsc.com.cn
	张岚	销售副总监	18601241803	zhanglan@swsc.com.cn	王一菲	销售经理	18040060359	wyf@swsc.com.cn
	杨薇	资深销售经理	15652285702	yangwei@swsc.com.cn	王宇飞	销售经理	18500981866	wangyuf@swsc.com
	姚航	高级销售经理	15652026677	yhang@swsc.com.cn	路漫天	销售经理	18610741553	lmtyf@swsc.com.cn
	胡青璇	高级销售经理	18800123955	hqx@swsc.com.cn	马冰竹	销售经理	13126590325	mbz@swsc.com.cn
广深	郑龔	广深销售负责人	18825189744	zhengyan@swsc.com.cn	丁凡	销售经理	15559989681	dingfyf@swsc.com.cn
	杨新意	广深销售联席负责人	17628609919	yxy@swsc.com.cn	陈紫琳	销售经理	13266723634	chzlyf@swsc.com.cn
	张文锋	高级销售经理	13642639789	zwf@swsc.com.cn	陈韵然	销售经理	18208801355	cyryf@swsc.com.cn
	龚之涵	销售经理	15808001926	gongzh@swsc.com.cn				