



第1篇
技术简史





第 1 章 10 点知识， 快速了解 Sora

学习提示

在数字时代的浪潮中，视频内容已成为信息传播和娱乐产业的核心驱动力。随着 AI（Artificial Intelligence，人工智能）技术的飞速发展，视频生成模型正逐渐从概念走向现实，其中 Sora 视频生成模型凭借其强大的技术实力，正引领着这一变革的浪潮。

本章围绕 Sora 的相关内容展开介绍，包括 Sora 的概念、特点、优势、用途及影响等，以便读者进一步了解 Sora。

1.1 认识什么是 Sora

Sora 是 OpenAI 公司推出的一款高质量的 AI 视频生成模型，引领了 AI 和视频创作领域的革命性飞跃。这款划时代的工具，借助先进的生成式 AI 技术，将文本描述转化为栩栩如生、充满创意的视频内容，如图 1-1 所示。



图 1-1 Sora 模型生成的视频内容

专家提醒

自 Sora 发布以来，科技界为之沸腾，它标志着内容创作的新纪元已然来临。OpenAI 公司通过 Sora 再次展现了其在 AI 领域的强大优势，不断拓展 AI 技术的边界。

Sora 以其独特的技术架构和高效的学习机制，在视频内容生成领域展现出了无与伦比的优势。Sora 不仅能够迅速捕捉和学习各种视频风格和特征，还能通过深度学习算法生成高质量、富有创意的视频内容。Sora 的出现，不仅极大丰富了视频创作手段，也为视频制作行业带来了前所未有的便利和效率提升。

Sora 的推出不仅是一项技术壮举，更是视频创作方式的一次颠覆性变革，它简化了视频创作流程，使制作高质量视频变得更加容易，让创作者、营销人员和教育工作者能以前所未有的便捷性和灵活性实现创意的落地。

Sora 的独到之处在于其核心功能——将文本描述转化为视频内容，这一功能使其从众多视频创作工具中脱颖而出。通过运用先进的 AI 技术，包括自然语言处理和生成算法，Sora 能够理解文本输入并将其呈现为动态的、视觉上很震撼的视频，相关示例如图 1-2 所示。这一功能不仅代表了生成式 AI 技术的巨大创新，还实现了传统视频制作方法难以企及的创意和效率水平。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

Several giant woolly mammoths approach treading through a snowy meadow, their long woolly fur lightly blows in the wind as they walk, snow covered trees and dramatic snow capped mountains in the distance, mid afternoon light with wispy clouds and a sun high in the distance creates a warm glow, the low camera view is stunning capturing the large furry mammal with beautiful photography, depth of field.

步骤 02 生成的视频效果：

这是 Sora 生成的一段冬日长毛猛犸象漫步的视频效果。



图 1-2 冬日长毛猛犸象漫步的壮丽景象

中文大意：几只巨型毛茸茸的猛犸象踏着雪地缓缓前行，它们长长的毛发在微风中轻轻飘动，远处是覆盖着雪的树木和壮丽的雪山，午后的阳光透过薄云洒下了温暖的光芒，低角度的摄像机视角令人惊叹，捕捉到了这些大型动物漫步的壮丽景象，景深感强。

Sora 会尝试捕捉提示词中描述的所有元素，并将其融合在一起，以创造出一个逼真的场景。除了主体的展现外，Sora 还会在背景中描绘出被雪覆盖的树木和白雪皑皑的山脉，以营造出一种宏伟而宁静的氛围。

这段视频描绘了一幅壮观的画面，展现了北方寒冷地区的雪原环境以及猛犸象的恢宏形象，具有多重视觉效果和情感共鸣，整个画面展现出了一种柔和而温暖的光影效果，为画面增添了一抹柔美的色彩，与雪地的明亮形成对比，使画面更加生动。在构图方面，采用了低角度的摄像机视角，捕捉到了猛犸象漫步的壮丽景象，同时通过景深感强的摄影手法，突出了动物主体，增强了画面的立体感和层次感。

从图 1-2 所示的画面中可以看到，无论是视频的真实性、时长、稳定性、连贯性、清晰度，还是对文本内容的深刻理解，Sora 都展现出了卓越的水平。过去，制作这样一段视频可能需要花费大量的时间和精力，从剧本创作到镜头设计，每一个步骤都烦琐而耗时。然而，现在仅需一段简短的文本描述，Sora 便能够轻松生成震撼人心的大场面，这无疑让相关从业者感到震惊和不安。

此外，Sora 的 AI 驱动方法提供了无与伦比的定制性和可扩展性，它能够根据文本描述生成独特且定制化的内容，实现更高层次的个性化，让每一个视频都独一无二。这一独特功能不仅彰显了 Sora 的技术实力，更突显了它在数字时代彻底改变我们创作和消费视频内容方式的巨大潜力。

专家提醒

与其他需要手动选择视觉效果、动画和特效的视频创作工具相比，Sora 的自动化特性显著节省了时间，降低了高质量视频制作的门槛，让创作者能够更加专注于故事的叙述，而非烦琐的视频制作技术细节。

1.2 了解 Sora 的 5 个特点

Sora 是一个革命性的 AI 视频生成工具，其功能之强大，足以颠覆传统的视频制作方式。那么，Sora 具体能做什么呢？下面简单介绍 Sora 的功能特点。

① Sora 的核心功能是将文本描述转化为生动的视频内容。用户只需通过文字描述自己的创意和想法，Sora 就能够将这些想法迅速转化为具有视觉吸引力和连贯性的视频。不论是复杂的场景构建还是多个角色的互动，甚至是细致入微的动作和背景描绘，Sora 都能够轻松应对，生成令人惊叹的视频作品，相关示例如图 1-3 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

An adorable happy otter confidently stands on a surfboard wearing a yellow lifejacket, riding along turquoise tropical waters near lush tropical islands, 3D digital render art style.

步骤 02 生成的视频效果：

这是 Sora 生成的一段水獭站在冲浪板上冲浪的视频效果。



图 1-3 一只可爱的水獭站在冲浪板上

中文大意：一只可爱的快乐水獭穿着黄色救生衣自信地站在冲浪板上，沿着郁郁葱葱的热带岛屿附近碧绿的热带水域骑行，3D 数字渲染艺术风格。

专家提醒

这段视频画面的提示词呈现了一个令人愉快而有趣的场景，以可爱的水獭为主角，通过色彩艳丽的场景和水獭自信的表现，营造了一种轻松、愉悦的氛围。

2 Sora 拥有卓越的自然语言理解能力。Sora 不仅能准确解析用户给出的文本提示，更能捕捉到其中的情感色彩和创意精髓，从而生成富含情感表达的视频。无论是欢快的节奏还是悲伤的氛围，Sora 都能够通过精准的角色表情，将情感完美传达。

3 Sora 还具备多镜头生成的能力。这意味着，在一个生成的视频中，Sora 可以巧妙地切换不同的镜头，创造出丰富的视觉体验。同时，它还能够保持角色和视觉风格的一致性，使整个视频作品呈现出高度的统一性和协调性。

4 Sora 可以从静态图像出发，生成动态的视频内容。只需提供一个现有的静态图像，Sora 就能够通过先进的图像处理技术，准确地动画化图像内容，让静态图像焕发出生命的活力。

5 Sora 具有视频扩展的功能。无论是想要延长现有视频的时长，还是想要填补视频中的缺失帧，Sora 都能够轻松胜任。Sora 能够通过分析和学习视频内容，生成与原始视频风格和内容相一致的扩展部分，使整个视频作品更加完整和连贯。

1.3 熟知 Sora 和 5 种模型的对比

Sora 通过其长达一分钟的视频生成、高度真实感的视频效果以及对文本描述的理解和执行能力，与其他 AI 视频生成工具相比，展现出了明显的优势和特点。表 1-1 所示为 Sora 和其他模型的能力对比。

通过深入比较 Sora 与其他视频生成模型的能力，可以清晰地揭示出 Sora 的独特优势和创新之处。当其他视频生成模型还在为保持单镜头的稳定性而努力时，Sora 已经实现了多镜头的无缝切换，这种切换不仅流畅自然，而且镜头间对象的连贯性和一致性效果也远胜于其他工具，真正实现了降维打击。

图 1-4 所示为 Sora 生成的视频效果，图 1-5 所示为 Runway 生成的视频效果，这两个视频使用了完全相同的提示词，但 Sora 在视频时长、提示词理解、视频质量、连贯性以及对现实世界物理规律的模拟能力方面均优于 Runway。

表 1-1 Sora 和其他模型的能力对比

能力分类	能力	Sora	其他模型
底层技术	架构	变换器	U-Net 为主
	驱动方式	数据	图片
对于真实世界的理解 / 模拟能力	世界理解能力	可理解世界知识	弱
	数字世界模拟	支持	不支持
	世界交互能力	支持	不支持
	3D 一致性	强	弱
	长期一致性	强	弱
	物体持久性 / 连续性	强	弱
	自然语言理解能力	强	一般
基于模拟的视频编辑能力	无缝连接视频	强	弱
	运动控制	提示词	提示词 + 运动控制工具
	视频到视频编辑	支持	部分
	扩展生成视频	前 / 后	后
外显视频基础属性	视频时长	60 秒	2 ~ 17 秒
	原生纵横比	支持	不支持
	清晰度	1080P	最高 4K

专家提醒

U-Net 是一种深度学习网络结构，主要用于图像分割等计算机视觉任务。U-Net 网络结构采用了编码器—解码器 (Encoder-Decoder) 的设计思想，其中编码器负责提取图像的特征，而解码器则负责根据这些特征进行像素级别的预测。

U-Net 网络结构的特点之一是它采用了跳跃连接 (Skip Connection)，将编码器的特征图与解码器的特征图进行连接，以便解码器能够利用编码器的低级特征进行更精确的预测。这种跳跃连接的设计使 U-Net 网络能够在保持高级语义特征的同时，不丢失低级细节信息，从而提高了图像分割的精度。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

Historical footage of California during the gold rush.

步骤 02 生成的视频效果：



这是 Sora 生成的一段淘金热期间加利福尼亚州的历史镜头，画面采用了特定的色调和滤镜来模拟 19 世纪淘金热时期的氛围，使用了暗淡的色彩和复古的色调来营造画面的历史感，展现出了当时加利福尼亚的野外景象，包括山脉、小溪、森林等自然景观，以及淘金者在这些景观中劳作的场景。



图 1-4 Sora 生成的视频效果

中文大意：淘金热期间加利福尼亚州的历史镜头。

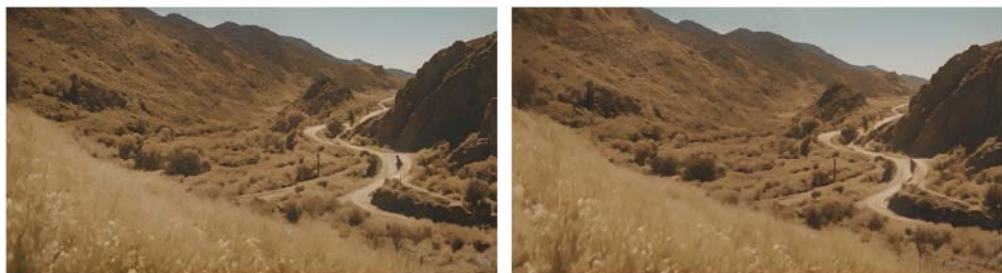


图 1-5 Runway 生成的视频效果



经过上述的细致对比，可以清晰地看出，Sora 不仅在整体上完全还原了提示词中描述的场景，而且在细节上也做得非常出色，特别是当时的城镇、商店等，充分展示当时社会的面貌和建筑风格，让人仿佛置身于真实世界之中。Sora 通过色调、景观、活动等方面展现出了 19 世纪加利福尼亚淘金热时期的历史场景和人文风貌。

相比之下，Runway 虽然基于 Stable Diffusion 技术，但受限于其模型训练的精度，生成的历史场景在细节上显得较为粗糙，尤其是淘金热期间加利福尼亚州的建筑场景也没有，更没有人们在淘金热期间的相关活动。

表 1-2 所示为主流的视频生成模型对比。通过与其他 5 种视频生成模型的对比，可以更加清晰地认识到 Sora 的独特之处和优势所在。无论是对于专业创作者还是普通用户来说，Sora 都是一个值得考虑和选择的 AI 视频生成工具。

表 1-2 主流的视频生成模型对比

模 型	开发团队	推出时间	是否 开源	外显视频基础属性		
				长 度	每秒帧数	分辨率 /PPI
Gen-2	Runway	2023 年 6 月	否	4 ~ 16 秒	24	768 × 448 1536 × 896 4096 × 2160
Pika	PIKA Labs	2023 年 11 月	否	3 ~ 7 秒	8 ~ 24	1280 × 720 2560 × 1440
Stable Video Diffusion	Stability AI	2023 年 11 月	是	2 ~ 4 秒	3 ~ 30	576 × 1024
Emu Video	Meta	2023 年 11 月	否	4 秒	16	512 × 512
W.A.L.T	谷歌	2023 年 12 月	否	3 秒	8	512 × 896
Sora	OpenAI	2024 年 2 月	否	60 秒	未知	最高 1080

专家提醒

上述视频生成模型的特点对比如下。

1 Gen-2 以出色的影视级构图和运镜能力著称，画面清晰度与精美度均达到了最高水平，其最新版本甚至可以生成 4K 画质的视频。

2 Pika 1.0 以其强大的语义理解能力脱颖而出，但在画面一致性方面还有一定的提升空间。

3 Stable Video Diffusion 作为第一个基于图像模型 Stable Diffusion 的生成式视频基础模型，它在视频生成领域具有里程碑意义。Stable Diffusion 是一种机器学习模型，该模型能够利用文本描述生成详细的图像，并可以用于图像修复、图像绘制、文本到图像和图像到图像等任务。

4 Emu Video 在视频生成质量和文本忠实度上表现出色，为用户提供了高质量的视频生成体验。

5 W.A.L.T 模型采用变换器+扩散的架构，旨在同时解决计算成本和数据集问题，为视频生成带来了更高效的解决方案。

6 Sora 模型同样采用变换器+扩散的架构，且在语义理解能力、复杂场景变化模拟能力以及一致性方面实现了突破性的进展，为用户提供了更加出色的视频生成效果。

1.4 知晓 Sora 的 5 个关键优势

当我们深入了解 Sora 这一视频生成模型时，不难发现其具备的多项核心优势，正是这些优势使 Sora 在视频生成领域脱颖而出，为用户提供了前所未有的视频生成体验。下面简单介绍 Sora 的核心优势。

1 Sora 以其高效快速的特点，赢得了用户的青睐。相较于传统的视频制作流程，Sora 能够根据用户提供的文字迅速生成视频，大大节省了制作时间和成本。

2 Sora 的高度定制化特性，为用户提供了更为广阔的创意空间。用户可以根据自己的需求，定制视频的内容和风格等，这使每一个生成的视频都充满了个性化和创意。无论是企业宣传、个人表达还是其他需求，Sora 都能满足用户的个性化需求。

3 Sora 的自动化程度也非常高。Sora 能够自动完成从文本到视频的转换，减少了人工干预和烦琐的操作。这意味着用户无需具备专业的视频制作技能，也能轻松生成高质量的视频，这一特性使 Sora 更加易于使用和普及。

4 Sora 生成的视频具有良好的跨平台兼容性。无论是在计算机、手机还是其他设备上，用户都能顺畅地播放 Sora 生成的视频，这种跨平台兼容性为用户提供

了更多的选择和便利。

5 Sora 的可扩展性也是其独特之处。随着技术的不断进步和应用场景的逐渐拓展，Sora 的功能和应用场景也将不断扩展和完善，这意味着 Sora 的未来充满了无限可能性和潜力。

1.5 掌握 Sora 的 7 个用途和使用范围

Sora 作为 OpenAI 公司推出的创新视频生成工具，为众多领域的应用提供了无限的可能性。无论是在娱乐与媒体、教育与培训、广告与营销、游戏开发、虚拟现实与增强现实，还是在艺术与文化创作，甚至是在个人创作与分享领域，Sora 都展现了其独特的魅力和巨大的潜力，相关介绍如下。

1 娱乐与媒体领域。Sora 以其出色的技术，为电影、电视节目和动画片注入了丰富的视觉效果和引人入胜的故事情节，极大地提升了作品的观赏性和吸引力，相关示例如图 1-6 所示。通过利用 Sora，制作团队能够迅速生成高质量的视频内容，显著缩短了生产周期并降低了成本。同时，Sora 也为社交媒体平台上的内容创作者提供了有力的支持，他们可以通过 Sora 轻松制作出生动、有趣的视频，从而增加在平台上的曝光度，吸引更多的关注和互动。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A cartoon kangaroo disco dances.

步骤 02 生成的视频效果：

这是 Sora 生成的一段袋鼠在舞台上跳舞的视频效果。



图 1-6 袋鼠在舞台上跳起了舞蹈



图 1-6 袋鼠在舞台上跳起了舞蹈（续）

中文大意：卡通形象的袋鼠跳起了迪斯科舞蹈。

② 教育和培训领域。Sora 所创造出的沉浸式和互动式学习环境，极大地激发了学生的学习兴趣 and 积极性，使教学质量和个性化程度得到了显著提升。教师可以利用 Sora 快速生成生动、直观的教学视频，为学生提供更加有趣和高效的学习材料。此外，Sora 还可以被广泛应用于培训材料的制作中，为学生提供更加便捷、高效的学习体验。

③ 广告与营销领域。企业可以利用 Sora 自动生成宣传或产品展示视频，提升品牌知名度和市场竞争力。Sora 的快速原型制作功能使营销团队能够迅速制作出创意十足的广告内容，吸引更多目标受众。同时，Sora 还能够通过生动、详细的视频展示产品的功能、使用场景和优势，帮助企业从市场中脱颖而出。

④ 虚拟现实与增强现实领域。Sora 以其独特的技术优势为虚拟现实（Virtual Reality, VR）和增强现实（Augmented Reality, AR）应用提供了丰富的动态内容支持。通过整合先进的图像处理技术和创新算法，Sora 能够生成高质量、逼真的虚拟场景和物体，为用户带来前所未有的沉浸式体验。无论是探索遥远的星球、漫步于古代的城市，还是与虚拟角色进行互动，Sora 都能为用户带来身临其境的感受。

⑤ 游戏开发领域。游戏开发者可以利用 Sora 制作游戏中的角色动画和场景效果，为游戏增添更高的交互性和趣味性。Sora 的创意应用为游戏开发带来了新的可能性和挑战，推动了游戏行业的创新和发展。

⑥ 艺术与文化创作领域。艺术家和文化创作者可以利用 Sora 创作出富有创意和表现力的视频艺术作品，从而推动数字艺术的发展和创新。无论是制作短片、音乐还是数字绘画，Sora 都能为创作者提供强大的技术支持。Sora 这种数字艺术的创作方式，不仅拓宽了艺术家的创作空间，还为观众带来了全新的艺术体验。

⑦ 个人创作与分享领域。Sora 为个人用户提供了便捷的工具。个人用户可以利用 Sora 进行创意视频制作并分享到社交媒体平台，展示自己的才华和创意，与他人分享自己的作品和想法。

1.6 发现文生视频模型 4 个火爆的原因

在 Sora 之前，市场上已经涌现出一批文生视频的平台和工具，其中 Pika 和 Runway 两家公司在 2023 年就已经推出了自己的文生视频模型。图 1-7 所示为 Pika 官方网站页面。然而，这些早期的模型所生成的视频，其主体往往只能进行缓慢的移动，且时长相对较短。下面讲解文生视频模型 4 个火爆的原因。

① Pika 作为一款文生视频平台，其功能远不止于此，除了基础的文生视频功能，即根据文本描述自动生成相应的视频内容外，Pika 还具备图生视频和视频生视频等多样化功能，如图 1-8 所示。

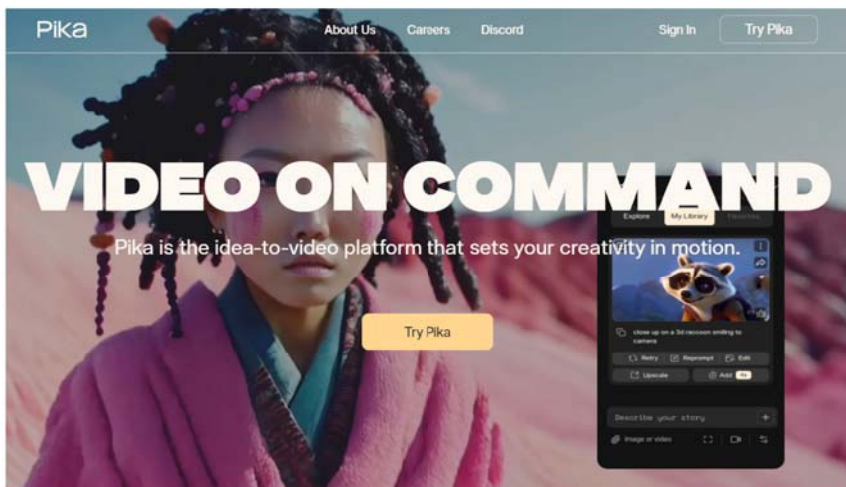


图 1-7 Pika 官方网站页面

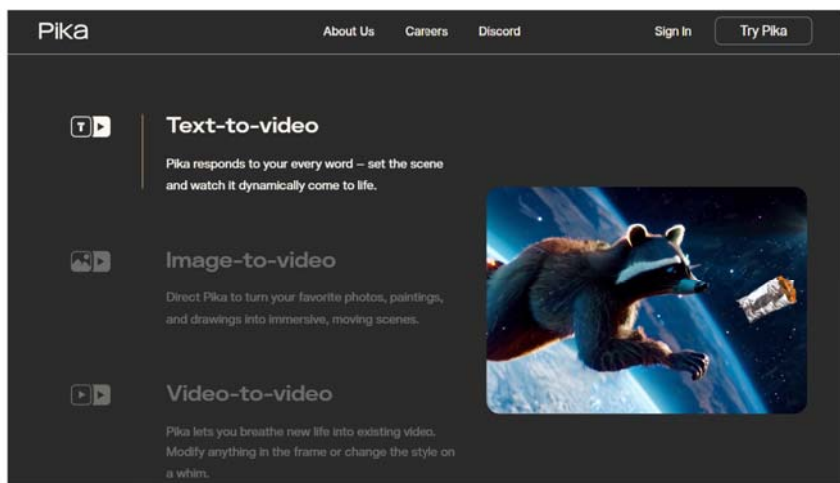


图 1-8 Pika 的文生视频功能演示

2 图生视频功能允许用户上传一张或多张图片，然后 Pika 根据这些图片中的内容，结合先进的图像处理 and 机器学习技术，自动生成一段连贯的视频，如图 1-9 所示。这为用户提供了一个全新的创作方式，使他们能够将静态的图片转化为动态的视频，进一步丰富了视觉体验。

3 视频生视频功能是 Pika 的另一个创新点，该功能允许用户上传一段已有的视频，然后 Pika 对这段视频进行深度分析，理解其中的内容、动作和场景等元素及其含义，基于这些信息生成全新的视频，如图 1-10 所示。视频生视频功能在视频编辑、内容创作和个性化推荐等领域有着广泛的应用前景。

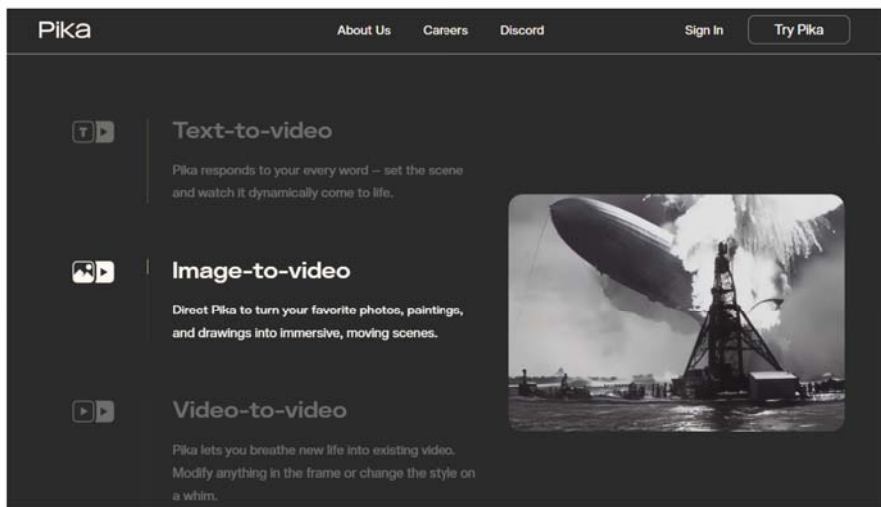


图 1-9 Pika 的图生视频功能演示

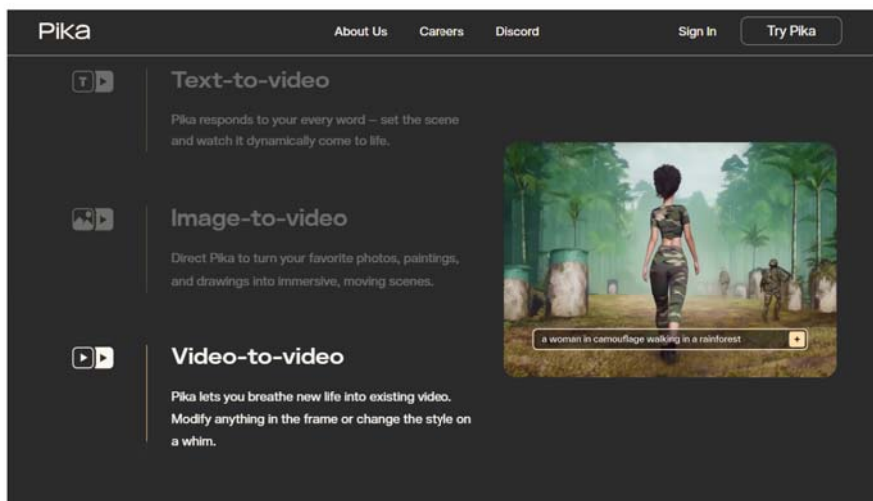


图 1-10 Pika 的视频生视频功能演示

④ Pika 还提供了丰富的编辑工具和特效库，用户可以根据自己的需求对生成的视频做进一步的编辑和美化，还可以修改视频的局部区域，如图 1-11 所示，Pika 使用户能够轻松打造出专业级的视频作品。

然而，尽管 Pika 已经具备了如此丰富的功能，但在 OpenAI 公司推出的 Sora 面前，其表现仍然显得有所不足。OpenAI 公司推出的 Sora 文生视频模型则展现出了更高的技术水平，其生成的视频不仅时长更长，而且主体运动更为流畅、逼真，仿佛赋予了视频主角以生命，相关示例如图 1-12 所示。这也进一步证明了 AI 技术的不断发展和创新，将为我们带来更多令人惊叹的应用和体验。

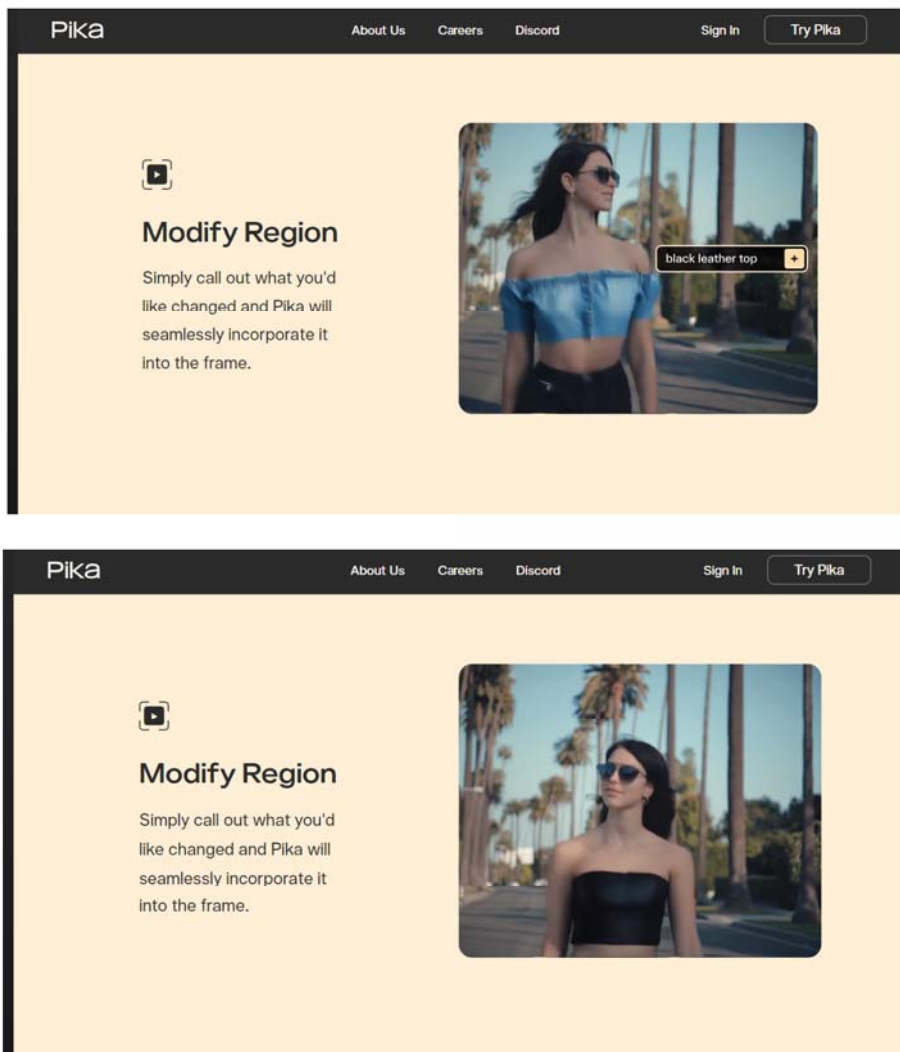


图 1-11 Pika 的视频编辑功能（修改区域）演示

▶ Sora 案例生成 ◀

步骤 01 输入的提示词:

An old man wearing blue jeans and a white T-shirt taking a pleasant stroll in Johannesburg South Africa during a beautiful sunset.

步骤 02 生成的视频效果:

这是 Sora 生成的一段老人在夕阳下散步的视频效果。



图 1-12 一个老人在夕阳下愉快地散步

中文大意：在一个傍晚美丽的夕阳下，一位身穿蓝色牛仔裤和白色 T 恤的老人，在南非约翰内斯堡愉快地散步。

从图 1-12 中可以看到，Sora 能够生成一个画面细腻、动态自然、背景丰富的视频，展现出在南非约翰内斯堡美丽的夕阳下，一名穿着蓝色牛仔裤和白色 T 恤的老人在小路上散步的场景。视频画面不仅富有生活气息，而且场景感非常强烈，能够给观众带来身临其境的感觉。

尽管这些对比都是基于 Sora 官方给出的示范效果进行的，但 OpenAI 公司作为一家在 AI 领域具有深厚积累的公司，其推出的产品和技术通常都备受关注。因此，有理由相信 Sora 在文生视频领域的表现确实达到了一个新的高度。

1.7 理解 Sora 给各行业带来的 3 个影响

Sora 的发布引起了全球范围内的广泛关注，人们纷纷惊叹于 AI 技术的飞速发展，而各大行业的精英人士和专家也纷纷发表了自己的看法。

在 Sora 发布后的数小时内，科技巨头埃隆·里夫·马斯克（Elon Reeve Musk）在社交媒体上回应了 gg（gg 是网络游戏用语 good games 的缩写，代指打得好，我认输的意思）humans（人类）的评论，他的意思是“人类输了”，他认为通过 AI 增强的人类将在未来几年创造出最好的作品，这一观点进一步证明了 Sora 在 AI 领域的极高地位以及其巨大的创作潜力。

甚至有专家预测，Sora 的出现可能意味着通用人工智能（Artificial General Intelligence, AGI）的实现时间将从 10 年缩短到 1 年。Sora 的技术特点在于它能够准确呈现视频细节，理解物体在现实世界中的存在状态，并生成具有丰富情感的角色，相关示例如图 1-13 所示。这是因为 Sora 学习大量视频，对世界的理解远远超过文字学习。这一发展趋势预示着，未来的 AI 将能够更深入地理解人类世界，从而推动各个领域的创新和发展。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A woman wearing a green dress and a sun hat, taking a pleasant stroll in Antarctica during a colorful festival.

步骤 02 生成的视频效果：

这是 Sora 生成的一个女人在南极洲愉快地漫步的视频效果。





图 1-13 一个女人在南极洲愉快地漫步

中文大意：一个穿着绿色连衣裙、戴着太阳帽的女人，在南极洲的节日中愉快地漫步。

从图 1-13 中可以看到，Sora 在生成视频时，能够精确地还原和展现细微的视觉元素，无论是场景的背景、物体的纹理还是角色的动作，Sora 都能以高度的真实感和清晰度来呈现。同时，Sora 能够识别并理解物体在现实世界中是如何存在、如

何与其他物体交互的，这种理解能力使它生成的视频更自然、流畅，更符合现实世界的逻辑。

然而，对于 Sora 的应用前景，也有业内人士持谨慎态度。他们认为，虽然 Sora 在技术上取得了突破，但要真正改变行业生态，还需要考虑行业规律和技术迭代的平衡。此外，生成式视频的信息量不如真实拍摄大，因此在对细节敏感的领域，如社交平台建设等，Sora 的应用可能还有一定局限性。总之，Sora 的发布不仅是一个技术里程碑，更是一个行业风向标，它象征着生成式 AI 大模型的热度与关注度将持续升温，并将为未来的科技产业带来更加深远的影响和变革。

无论你是否愿意承认，Sora 等 AI 技术正在与我们逐渐建立起更加紧密的联系。它们不仅改变了人们的工作方式，更在某种程度上重塑了人们的生活方式。只有积极拥抱这些变化，才能在这个日新月异的时代中保持竞争力。下面简单分析 Sora 给我们带来的影响。

① AI 改变工作和生活方式。 AI 正在深刻改变着人们的工作和生活方式。许多传统行业和岗位正在逐渐被 AI 所取代，如程序员的大量复制粘贴工作现在可由 ChatGPT 轻松完成，且生成的代码更加规范。同样，曾经需要庞大团队完成的任务，现在可能只需少数几人就能完成。

② 技术变革的双刃剑效应。 一方面，我们期待着新技术如 Sora 带来的应用前景和便利；另一方面，也有人担忧这些技术可能会抢走传统职业的饭碗，这种担忧并非无的放矢，因为技术的快速发展确实会对某些行业产生影响。其中，影视行业的从业者可能是最容易受到影响的群体之一。

图 1-14 所示为使用 Sora 生成的科幻电影片段。随着 Sora 的出现，它能够自动或半自动地生成视频，这可能会减少传统视频制作和编辑职位的需求。同时，对于影视行业的从业者来说，他们需要不断提升自己的技能和能力，以适应这一变革。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A close up view of a glass sphere that has a zen garden within it. There is a small dwarf in the sphere who is raking the zen garden and creating patterns in the sand.

步骤 02 生成的视频效果：

这是 Sora 生成的一段玻璃球体里的小矮人的视频效果。





图 1-14 玻璃球体里的小矮人

中文大意：一个玻璃球体的特写，里面有一个禅意花园。球体里有一个小矮人，他正在用耙子整理禅园中的沙子，并在沙子上创造图案。

③ 失业潮未必会发生。虽然 Sora 等新技术可能会对传统职业造成一定冲击，但并不意味着一定会引发失业潮。相反，随着技术的普及和应用，它可能会催生新的职业岗位和就业机会。此外，人类的创造力和智慧是技术无法替代的，因此我们应该积极面对技术变革带来的机遇和挑战，努力提升自己的能力和素质，以适应未来社会的发展需求。

1.8 领悟 Sora 给 AI 行业带来的 3 个进步

我们与 Sora 有何关系？这是许多人在 2023 年面对生成式大模型（如 ChatGPT）的崛起时所思考的问题。随着文生图等技术的日益成熟，文生视频已崭露头角，成为多模态大模型发展的下一个重要方向。展望未来，行业专家普遍认为，

2024 年大模型领域的竞争将更加激烈，而多模态大模型将引领生成式 AI 的新潮流，推动整个 AI 行业的进步，具体表现在以下 3 个方面。

① AI 文生视频技术推动短剧市场变革。AI 文生视频作为多模态应用的下一个重要领域，其根据文字提示直接生成视频的能力，预示着短视频市场即将迎来重大变革。这一技术有望显著降低短视频制作成本，从而解决“重制作而轻创作”的问题，使短视频制作的重心回归到高质量的剧本创作上。

② 多模态大模型算法突破对科技产业的影响和改变。多模态大模型算法的突破将为自动驾驶、机器人等技术带来革命性的进步，同时生成式 AI 对科技产业将产生长期影响，建议大家多关注算力、算法、数据、应用等环节的龙头企业。

③ 多模态在 AI 商业应用中的重要作用和潜力。多模态是 AI 商业应用的重要起点，有望为企业带来真正的降本增效效果。企业可以利用节省下来的成本提升产品、服务质量或进行技术创新，从而推动生产力的进一步提升。此外，多模态的发展还可能催生新的、更广阔的用户生成内容平台，为整个行业带来更大的发展空间。

1.9 面对 Sora 到来的 6 个应对方法

面对 Sora 的崛起，我们应积极适应并善加利用，以开放、审慎和批判性的态度来应对。在充分利用 Sora 带来的机遇的同时，也要关注其可能带来的风险和挑战，并努力寻求平衡和可持续发展的道路，相关方法如下。

① 明确使用界限。我们应遵循科技伦理规范，不利用 Sora 进行不道德或欺诈性行为。虽然中华人民共和国科学技术部监督司印发的《负责任研究行为规范指引（2023）》并未直接禁止使用 Sora 等生成式 AI 模型，但强调了科研伦理的重要性。这意味着，我们不能利用 Sora 等 AI 模型进行不道德或欺诈性行为，如直接生成申报材料或将其列为成果共同完成人。

② 利用 Sora 开阔视野、促进学习效果。通过 AI 大模型，我们能够轻松获取各个领域的知识，打破曾经的信息壁垒。现在，学习已经变得前所未有的容易，只要我们有意愿，就能随时随地进行学习。图 1-15 所示为 AI 大模型 ChatGPT 的使用示例，可以让它扮演老师来传授相关方面的知识。

③ 转变观念和思维方式。我们要与时俱进地调整工作、教育和终身学习观念，以适应 AI 时代的需求。无论是工作观念、对孩子的教育观念，还是自己的终身学习观念，都需要与时俱进。在 AI 快速发展的今天，我们应教会孩子学习方法，让他们学会适应社会变化，更要培养他们在未知的未来中迎接挑战的能力。

④ 保持审慎态度。避免盲目崇拜和过度依赖 AI，保持对其的理性和客观认识。虽然 Sora 等 AI 大模型为我们带来了前所未有的便利和机遇，但我们也必须认识到它们并不是万能的。在追求科技进步的同时，仍须保持对 AI 的理性和客观认识，避免盲目崇拜和过度依赖。

ChatGPT

人工智能 (AI) 有许多应用领域, 主要包括但不限于以下几个方面:

1. **医疗健康:** AI在医疗诊断、治疗计划、药物研发、个性化医疗等方面发挥重要作用。比如通过深度学习分析医学影像, 辅助诊断疾病。
2. **自动驾驶:** AI技术使得汽车能够实现自动驾驶, 通过分析环境数据来做出行驶决策。
3. **金融服务:** 在风险评估、交易、个人理财等领域, AI帮助提高决策质量和效率。
4. **制造业:** AI在生产流程优化、质量控制、设备维护预测等方面应用广泛。
5. **零售业:** 利用AI进行库存管理、个性化推荐、顾客行为分析等。
6. **语言处理:** 在翻译、语音识别、自然语言理解等方面, AI技术已取得显著成就。
7. **教育:** AI可以提供个性化学习体验, 辅助教师制定教学计划。
8. **游戏:** 在游戏设计、玩家行为分析、实时交互等方面, AI技术也在不断发展。
9. **安全:** AI在网络安全、公共安全等领域用于威胁检测和响应。
10. **艺术和创意产业:** AI在音乐创作、图形设计、内容生成等方面也有应用。

图 1-15 AI 大模型 ChatGPT 的使用示例

5 关注风险和挑战。我们应该关注 AI 可能带来的风险和挑战。例如, 数据隐私和安全性、AI 的决策透明度和公平性等问题都需要我们深入思考和解决。因此, 在利用 Sora 等 AI 大模型的同时, 也应该加强对其的监管和规范, 确保其发展符合社会的公共利益和伦理原则。

6 培养批判性思维能力。我们应该注重培养自己的批判性思维能力, 在面对 AI 生成的信息和知识时, 我们应该保持独立思考和判断能力, 不盲目接受和传播未经证实的信息。同时, 我们也应该学会识别和评估 AI 生成结果的可靠性和准确性, 以便更好地利用它们来指导我们的决策和行动。

1.10 抓住 Sora 风口的 6 个准备工作

在 AI 领域的热潮中, Sora 模型成为众人瞩目的焦点。面对 Sora 这样的文生视频模型带来的技术革新, 普通人又该如何把握其带来的机遇呢? 首先, 我们得承认 Sora 模型所带来的影响, 它生成的视频效果让人惊叹不已。这里为大家展示一段 Sora 的视频生成提示词, 以激发我们的想象力, 相关示例如图 1-16 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词:

Aerial view of Santorini during the blue hour, showcasing the stunning architecture of white Cycladic buildings with blue domes. The caldera views are breathtaking, and the lighting creates a beautiful, serene atmosphere.

步骤 02 生成的视频效果：

这是 Sora 生成的一段圣托里尼岛的鸟瞰视频效果。



图 1-16 圣托里尼岛的鸟瞰视频

中文大意：蓝色时刻圣托里尼岛的鸟瞰图，展示了带有蓝色圆顶的白色基克拉迪建筑的令人惊叹的建筑风格。火山口的景色令人叹为观止，灯光营造出了美丽、宁静的氛围。

这段视频展示了带有蓝色圆顶的白色基克拉迪建筑的建筑风格，营造出了美丽、宁静的氛围。即使没有看过这段视频，这段提示词描述的场景也足以让我们脑海中浮现出一个带有蓝色圆顶的白色基克拉迪建筑的画面。而如果我们只是简单地写下“蓝色时刻圣托里尼岛的鸟瞰图”，那么生成的视频效果可能会大打折扣。

从本示例的视频中可以看到，Sora 展现出的画质和流畅度，让人不禁感叹：“这真的是 AI 做的吗？”是的，这确实是 AI 技术的力量。然而，对于普通人来说，AI 技术的神秘和高深莫测往往伴随着一种期待和焦虑。期待的是能够借助 AI 技术为自己的业务带来质的飞跃，焦虑的是不知道如何融入这场技术革命。

面对这样的焦虑，首先要冷静下来。AI 技术虽然强大，但它始终是一个工具，真正能够创造价值的是我们对业务的深入理解。因此，普通人在开始使用 AI 视频时，不必盲目跟风，而是要结合自己的业务，思考 AI 技术如何为自己的业务带来增值。另外，还要明确一点，目前 OpenAI 公司只是发布了演示视频和一篇研究论文，Sora 真正的技术应用还未全面开放，所以我们需要保持冷静和理性。

Sora 之所以如此惊艳，并不是因为它从零开始原创了一个模型，而是站在了 OpenAI 公司其他成功产品的肩膀上，借鉴了如 ChatGPT 等大语言模型的思路和 OpenAI 公司内部的成功经验，同时还付出了巨大的模型训练成本，这也是其他公司难以复制 Sora 的原因之一。对于普通人来说，想要把握住 Sora 带来的机遇，可以提前做好以下几点准备。

1 关注技术动态。时刻关注 AI 领域的技术发展，了解最新的技术动态和趋势，特别是与视频生成相关的技术。这样，当新的技术出现时，就能够迅速发现其背后的价值和意义，以便在机会来临时迅速把握住。

2 结合业务思考。将 AI 技术与自己的业务相结合，思考如何利用 AI 技术解决业务中的痛点，提升工作效率和质量。在 AI 浪潮中，不要盲目跟风，而是要根据自己的实际情况和需求，做出最适合自己的选择。

3 深入学习内容创作。虽然 AI 技术能够高效快捷地生成视频内容，但内容的本质仍然需要人来把握。因此，需要深入学习内容创作的底层逻辑，理解如何让作品吸引更多观众、更好地传递价值，这是用好 AI 技术的前提。

4 培养创新思维。AI 技术的发展为内容创作带来了无限的可能性，普通人可以发挥自己的想象力，结合 AI 技术创作出更具创意和个性化的视频内容。

5 建立自己的素材库。为了丰富自己的创作灵感和提高视频制作能力，应该积极建立自己的素材库。一个有效的方法是多观看一些经典电影或剧集，并从中选择精彩的片段或画面进行截屏或保存。这样，就可以逐渐积累起一个丰富多样的素材库，更加得心应手地创作出富有创意和吸引力的视频作品。

6 提升视频制作技能。如果你有余力，学习一些视频剪辑和构图的入门技巧将是非常有益的。掌握基础的摄影摄像知识，并熟悉一些视频剪辑软件的操作，将使你能够更自如地指导 AI 生成视频，这些技能不仅有助于提升视频质量，还能让你在创作过程中更加游刃有余。

总之，面对 Sora 这样的技术革新，普通人既要有期待和热情，也要保持冷静和理性。可以通过关注技术动态、结合业务思考和培养创新思维等，提前做好准备，抓住这一技术革新的机遇。

本章小结

本章主要介绍了 Sora 的相关基础知识，包括 Sora 的基本概念、特点、与其他模型的对比、关键优势、用途和使用范围等，并对 Sora 给各行业带来的影响与进步进行了详细讲解，最后介绍了面对 Sora 到来的 6 个应对方法，帮助大家面对 Sora 带来的机遇，提前做好相关的准备工作。

课后习题

鉴于本章知识的重要性，为了让大家能够更好地掌握所学知识，下面将通过课后习题进行简单的知识回顾和补充。

- (1) 请根据你的理解，对 Sora 的概念和特点进行相关讲解。
- (2) 请对 Sora 的关键优势进行简单概括。



课后习题 1



课后习题 2



第 2 章 13 个要点， 掌握 Sora 创新能力

学习提示

随着数字化时代的到来，视频已成为人们获取信息和娱乐的重要方式。然而，高效的视频生成技术一直是业界的挑战和追求。在这一背景下，Sora 的出现为我们提供了一种全新的解决方案。那么，Sora 究竟是如何生成视频的？本章将深入解析 Sora 的创新能力，揭示其高效生成视频的秘密。

2.1 生成长达 60 秒的超长视频

利用 Sora 可以生成长达 60 秒的超长视频，为视频制作带来了许多新的可能性和更多的选择，通过这种技术制作的视频特点主要体现在以下几个方面，如图 2-1 所示。

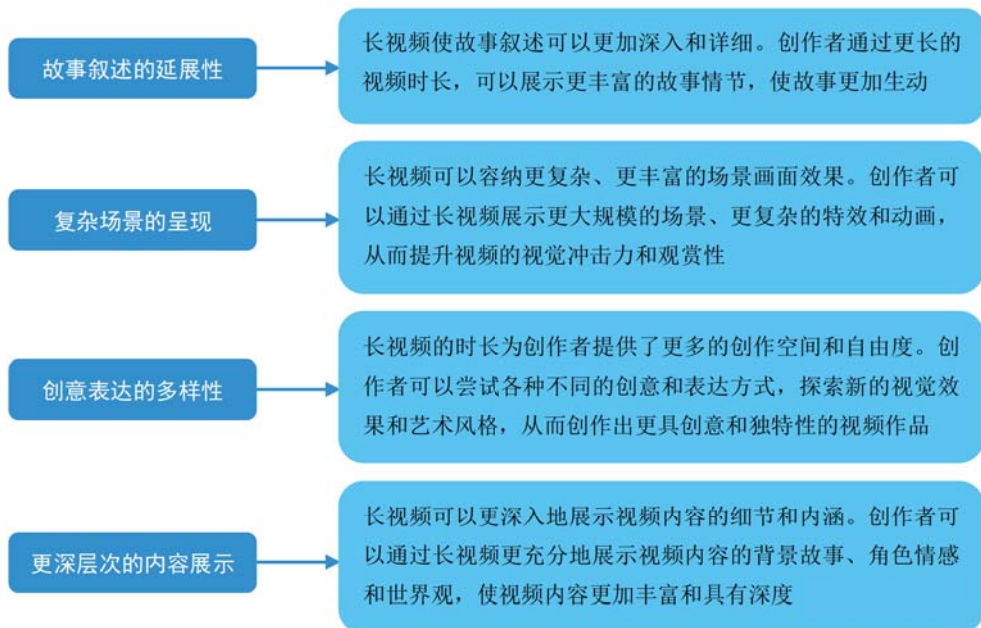


图 2-1 Sora 生成的视频特点

综上所述，Sora 生成长达 60 秒的超长视频为创作者提供了更广阔的创作空间和更丰富的创作选择，助力他们创作出更具吸引力的视频，相关示例如图 2-2 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

步骤 02 生成的视频效果：

这是 Sora 生成的一段时尚女士穿过东京街道的视频效果，视频时长达 59 秒。



图 2-2 时尚女士穿过东京街道视频

中文大意：一位时尚的女士穿过东京街道，街道上温暖的霓虹灯和动态城市标志闪烁着。她穿着一件黑色皮夹克，一条长长的红色连衣裙，黑色靴子，手提一只黑色的手袋。她戴着墨镜，涂着红色的口红，自信而悠闲地走着。街道潮湿而反光，形成了色彩灯光的镜面效果。许多行人来来往往。

通过欣赏上面这段 AI 视频，可以得到以下几点感受。

- ① 场景充满时尚和现代感，通过强烈的色彩对比和环境细节的描绘，创造出了一种电影般的视觉效果。
- ② 街道上温暖的霓虹灯和动态城市标志色彩饱和度高、变化频繁，这种光影效果给画面增添了许多动感和活力。
- ③ 视频画面的播放十分流畅，女主角自信而悠闲地走着，个性化的服装增强了她的个人魅力和气场，使画面更加生动。

④ 看到这样的视频画面，会让人有一种错觉，就是不相信这些视频画面是由 AI 生成的，以为这是一段实拍的真实画面。

2.2 生成的视频具有丰富的细节

Sora 能够生成高质量的视频内容，包括丰富的细节、逼真的场景和人物，以及自然、流畅的动作和过渡效果，这在视频生成领域是一项重大创新。

Sora 实现高质量的视频生成主要依靠深度学习技术、生成式对抗网络 (Generative Adversarial Network, GAN) 技术以及循环神经网络技术，这些技术的结合使 Sora 能够生成具有逼真感和高质量的视频内容，为用户提供了全新的视频创作体验，相关示例如图 2-3 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

The Glenfinnan Viaduct is a historic railway bridge in Scotland, UK, that crosses over the west highland line between the towns of Mallaig and Fort William. It is a stunning sight as a steam train leaves the bridge, traveling over the arch-covered viaduct. The landscape is dotted with lush greenery and rocky mountains, creating a picturesque backdrop for the train journey. The sky is blue and the sun is shining, making for a beautiful day to explore this majestic spot.

步骤 02 生成的视频效果：

这是 Sora 生成的一段蒸汽火车行驶在高架桥上的视频效果。



图 2-3 蒸汽火车行驶在高架桥上的视频



图 2-3 蒸汽火车行驶在高架桥上的视频（续）

中文大意：格伦芬南高架桥是英国苏格兰的一座历史悠久的铁路桥，横跨马莱格镇和威廉堡镇之间的西高地线。当一列蒸汽火车离开大桥，在拱形高架桥上行驶时，这是一个令人惊叹的景象。风景中点缀着郁郁葱葱的绿色植物和岩石山脉，为火车之旅创造了风景如画的背景。天空湛蓝，阳光灿烂，这是探索这一雄伟景点的美好一天。

下面对 Sora 生成的这段视频效果进行相关分析。

① 景点特色。视频画面展现了格伦芬南高架桥的独特之处，特别是蒸汽火车在高架桥上驶过的场景，这种效果在视觉上非常引人注目，吸引着观众的注意力。

② 自然环境。画面中的背景呈现了苏格兰典型的自然风光，包括郁郁葱葱的绿地和崎岖的山脉，这些元素共同构成了一幅宜人的自然画卷，为视频增色不少。

③ 天气状况。描述中提到天空湛蓝、阳光灿烂，这一点也在画面中得到了体现，良好的天气为整个场景提供了明亮的光线，使画面更加清晰、生动。

④ 火车运行。视频中的蒸汽火车在高架桥上行驶，展示了其独特的设计和工艺，同时也传达了城市的历史感，蒸汽火车的行驶为整个画面注入了动感，吸引了观众的视线。

总体而言，视频通过格伦芬南高架桥的独特场景、自然环境以及蒸汽火车的行驶，成功地呈现出了一幅美丽而宏伟的画面，视频效果具有丰富的细节与逼真感，让观众能够感受到苏格兰乡村的迷人之处。

2.3 具有准确理解自然语言的能力

Sora 拥有深入的语言理解能力，它能够准确理解用户提供的语言提示，并根据这些提示生成具有丰富情感的角色，这个功能的实现主要基于自然语言理解 (Natural

Language Understanding, NLU) 和深度学习技术, 以及应用了 DALL·E 3 中引入的重新字幕技术到视频。

Sora 可以分析和理解提示词中的语义、情感和意图, 从而准确把握用户的需求和要求。下面分析 Sora 在生成角色方面对自然语言的理解, 如图 2-4 所示。

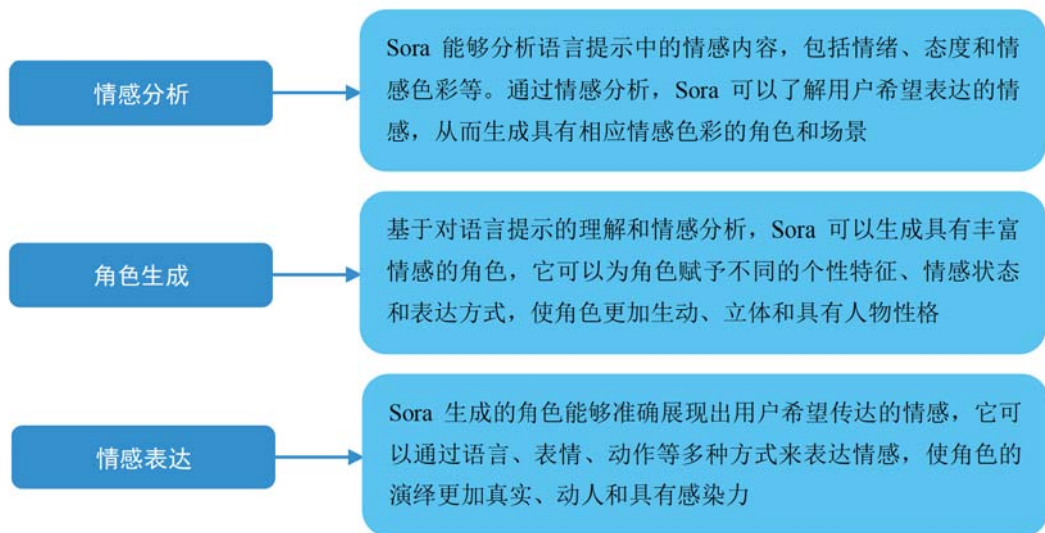


图 2-4 Sora 在生成角色方面对自然语言的理解

2.4 具有较高的图像和视频分辨率

Sora 可以生成高分辨率的图像与视频效果, 图像分辨率可以达到 2048×2048 像素, 具有更高的分辨率和更大的像素密度。这种图像效果适用于多种应用场景, 包括印刷、数字艺术、网络图片等, 能够呈现出更多的细节和更高的图像质量。

视频分辨率可以达到 1920×1080 或 1080×1920 像素, 分别对应着横向和纵向的高清视频。这对于视频创作非常重要, 因为高分辨率的视频可以提供更清晰、更细腻的画面细节, 从而提升用户的观看体验。这种视频适用于多种播放场景, 如电视机、计算机屏幕、移动设备等, 能够提供良好的观看体验, 相关分析如图 2-5 所示。

专家提醒

值得一提的是, Sora 不仅在视频格式的多样性上表现出色, 更在内容生成流程上实现了简化。Sora 允许用户在较低分辨率下快速显示原型内容, 以便进行初步预览和调试。一旦确认无误, 用户可以直接在 Sora 的全分辨率模式下进行最终生成。这一特点不仅提高了内容创作的灵活性和效率, 还大大简化了视频内容的生成流程, 为用户带来了前所未有的便捷体验。

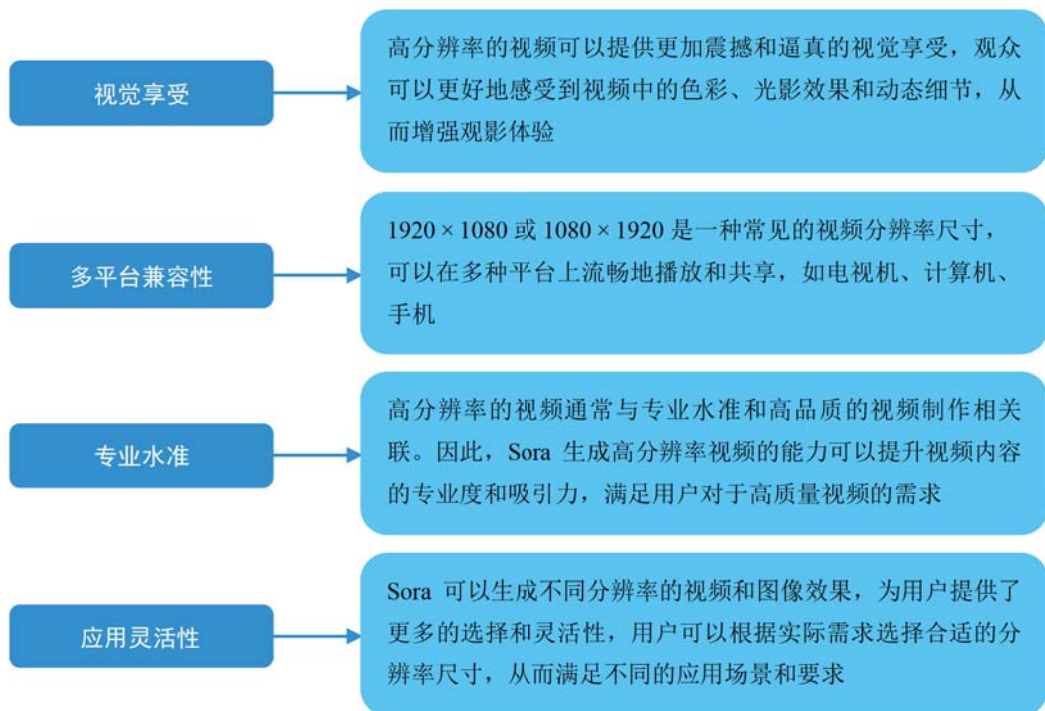


图 2-5 关于分辨率的相关分析

2.5 具有灵活的视频采样能力

Sora 模型在视频生成方面具有灵活的采样能力，可以生成不同宽高比的视频，包括竖屏 9 : 16、正方形 1 : 1 以及宽屏 16 : 9 这 3 种不同的视频尺寸，以及介于两者之间的所有尺寸，完美展现了 Sora 模型灵活的视频采样能力。

Sora 的视频采样灵活性得益于其强大的技术实力和算法优化，它能够自动适应不同视频格式的采样需求，无需手动调整参数或转换格式，从而大大简化了视频内容生成的流程。此外，Sora 还采用了高效的计算架构和数据处理方式，确保在采样过程中不会损失任何图像像素或细节。这些技术特点使 Sora 在视频生成领域具有极高的竞争力和市场价值，相关示例如图 2-6 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A turtle happily swims around in the water.

步骤 02 生成的视频效果：

这是 Sora 使用同一提示词生成的 3 段不同尺寸的视频效果，分别为竖屏 9 : 16、正方形 1 : 1 以及宽屏 16 : 9 的视频效果。选择哪种视频尺寸，通常取决于视频内容的用途、目标受众和观看平台。在移动设备上，竖屏和正方形可能更为流行，而在电视和电影领域，宽屏更为常见。



竖屏 9 : 16 的视频效果



正方形 1 : 1 的视频效果

图 2-6 3 段不同尺寸的视频效果





宽屏 16 : 9 的视频效果

图 2-6 3 段不同尺寸的视频效果 (续)



中文大意：一只乌龟在水里愉快地游来游去。

2.6 精进的画面构图和布局

Sora 团队研究发现，在原始横纵比的视频上进行训练，可以提高画面的构图和组合效果。他们将 Sora 与其他模型进行了比较，该模型将所有的训练视频素材都裁剪为正方形，这是训练图像或视频生成模型时的常见做法。结果发现，训练素材被裁剪成正方形的模型有时会生成只显示部分主体的视频。

相比之下，Sora 生成的视频在构图方面有所改进。通过对比可以看出，Sora 在视频生成方面的优势不仅在于其强大的技术实力和算法优化，更在于其对原始视频横纵比的充分利用和考虑。这种考虑使 Sora 在生成视频时，能够更好地保留原始视频的画面构图，从而让用户获得更加真实、生动的观看体验，相关示例如图 2-7 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A video footage of a person diving underwater.

步骤 02 生成的视频效果：

这是 Sora 生成的一段人物在海底进行潜水的视频效果，左图是训练素材被裁

剪成正方形的模型生成的效果，仅展示了部分主体内容；右图是 Sora 生成的视频效果，在取景和构图方面效果更好。

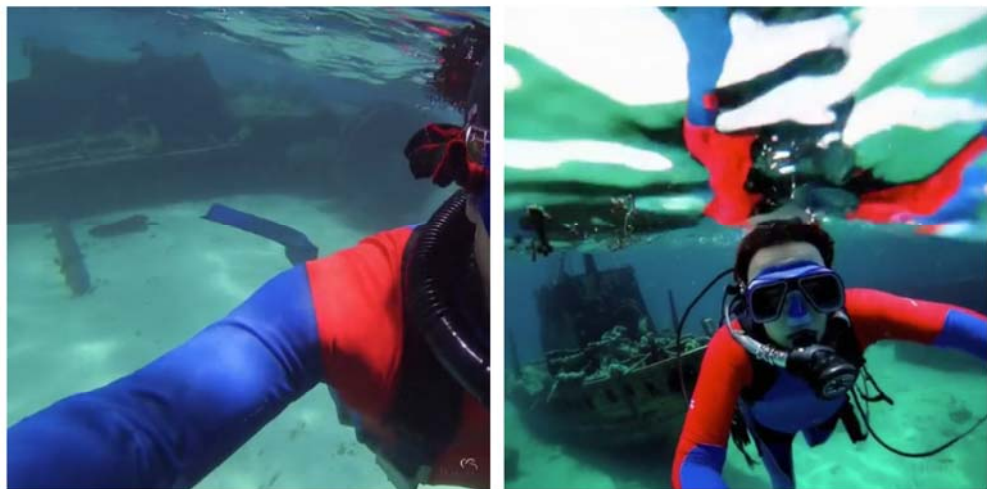


图 2-7 一个人在海底进行潜水视频

中文大意：一个人在海底进行潜水的视频画面。

2.7 保持视频画面的 3D 一致性

在 Sora 团队深入研究视频模型的过程中，一个引人注目的现象逐渐浮现出来：当进行大规模训练时，这些模型展现出了许多令人惊叹的“涌现”能力。这些“涌现”能力不仅令人印象深刻，更重要的是，它们赋予了 Sora 独特的视频生成能力，使其能够精确地模拟物理世界中的人、动物以及环境。

这些“涌现”能力的属性并非基于任何特定的归纳偏差 (Inductive Bias)，如 3D (Three-Dimensional，三维) 结构或物体识别等。相反，它们纯粹是模型在处理大量数据时自然产生的尺度现象。换句话说，这些属性是模型在庞大的数据集上进行训练时自我学习和自我优化的结果，而非人为预设或强加的。

这种无偏差的“涌现”能力，使 Sora 在模拟现实世界时更加灵活和真实。无论是模拟人物的动态行为、动物的奔跑跳跃，还是重现复杂的环境变化，Sora 都能够凭借其强大的“涌现”能力，呈现出令人信服的结果。

其中，3D 一致性就是 Sora “涌现”能力中的一个重要特点，使 Sora 可以生成具有镜头运动效果的动态视频，随着镜头的移动和旋转，人和场景元素在三维空间中始终会保持一致的运动，相关示例如图 2-8 所示。

📢 专家提醒

归纳偏差是指在机器学习算法中，模型对特定类型的数据或假设的偏好。这种偏好可能会导致模型在训练过程中偏向于某些解决方案，而忽略其他可能的、同样有效的解决方案。归纳偏差通常是由于模型的设计、参数的选择、训练数据的特性等因素引起的。

在视频模型中，归纳偏差可能表现为模型对某些类型的视频或场景有更强的识别能力，而对其他类型的数据则表现较差。例如，一个模型可能被设计为更擅长识别静态图像，而对动态视频的处理能力较弱。这种偏差可能会导致模型在处理复杂或多样化的视频数据时表现不佳，因为它可能过于依赖某些特定的特征或模式。

为了减轻归纳偏差的影响，研究人员通常会尝试不同的模型结构、训练策略或数据增强技术，以增加模型的泛化能力和适应性。这样可以帮助模型更好地处理各种类型的数据，提高其在不同场景下的性能表现。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

Beautiful, snowy Tokyo city is bustling. The camera moves through the bustling city street, following several people enjoying the beautiful snowy weather and shopping at nearby stalls. Gorgeous sakura petals are flying through the wind along with snowflakes.

步骤 02 生成的视频效果：

这是 Sora 生成的一段东京城熙熙攘攘的视频画面。

3D 一致性是视频生成中一个重要的概念，从图 2-8 中可以看到，它保证了 Sora 生成的视频在空间上具有连贯性和真实性。当镜头跟随人物向前推进时，Sora 能够精确地模拟和渲染出周围环境的细节和变化。



📢 专家提醒

Sora 具备出色的物理模拟能力，能够模拟真实世界中的物理规律，如重力、碰撞和摩擦力等，使生成的视频内容在动态表现上更加自然、真实。无论是风吹草动还是水流潺潺，Sora 都能够精准地模拟出这些自然现象，为用户带来身临其境的感受。



图 2-8 白雪皑皑的东京城市

中文大意: 美丽、白雪皑皑的东京城熙熙攘攘。镜头穿梭于熙熙攘攘的城市街道，跟随享受美丽的雪天并在附近的摊位购物的几个人。绚丽的樱花花瓣随着雪花在风中飘扬。

这意味着，无论是人物的行走、跑步或者跳跃，还是场景中的建筑物、树木等元素的移动，都能够与镜头的运动保持协调，呈现出更加真实和自然的视觉效果。因此，3D 一致性不仅增强了视频的视觉效果，还提升了用户观看体验。

2.8 保持主体和场景的一致性

以人物、动物或者物体为例，即使在它们被遮挡或离开画面的情况下，Sora 模型也能通过其强大的处理能力使它们以某种方式在视频中持续存在，相关示例如

图 2-9 所示。这种长期一致性（指 AI 视频的长程连贯性和物体永久性）的特性，能使 Sora 生成的视频更加自然、真实，给观众带来更好的观看体验。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A beautiful silhouette animation shows a wolf howling at the moon, feeling lonely, until it finds its pack.

步骤 02 生成的视频效果：

这是 Sora 生成的一段狼对着月亮嚎叫的剪影视频效果，即使在不同的视频和镜头部分，狼的外观、树的剪影效果以及视频的颜色风格也能够保持一致性。



图 2-9 狼对着月亮嚎叫的剪影

中文大意：这是一段美丽的剪影动画，展现了一只狼对着月亮嚎叫，感到孤独，直到它找到了自己的狼群。

长期一致性一直是 AI 视频生成领域面临的重要挑战。在采样长时间视频时，保持内容在时间上的连贯性对于视频生成模型来说尤为困难。然而，视频生成模型 Sora 在这方面表现出了不俗的能力。尽管并非在所有情况下都能完美应对，但 Sora 通常能够有效地处理短期和长期的依赖关系，确保生成的视频在内容上具有长程连贯性。

例如，在一段视频中，若一个角色在开始时身着红衣，那么不论视频如何切换场景或角度，该角色的红衣着装都将始终如一，保持高度的一致性。同样地，当视频描述一个人物从一张桌子移动到另一张桌子的过程时，Sora 的强大能力就得以凸显。即便视角发生转换，或是场景有所变换，人物与两张桌子之间的相对位置及其互动细节，都将被精准地维持和呈现。这种长期一致性的保持，不仅体现了 Sora 在视频处理上的深厚实力，也为观众带来了更为真实和沉浸的观影体验。

2.9 能够“与世界进行互动”

“与世界进行互动”是 AI 领域一个具有挑战性的目标，而 Sora 在这方面展现出了不俗的能力。在某些情况下，Sora 能够模拟出对世界状态产生影响的简单动作，使虚拟世界中的物体和角色能够与现实世界一样进行交互，使生成的视频内容更加真实、生动和富有情感，从而让生成的视频更具有情感共鸣和生动性，为用户带来全新的视频创作和观影体验。

图 2-10 所示的视频展示了一个画家站在画布前，用笔在画布上作画。这时，画家在画布上留下了新的笔触、新的颜色、新的线条以及新的形状，观众可以看到画家的手臂和手部动作，以及笔触在画布上的变化，这样的呈现方式使观众可以体验到画家创作的过程，并感受到画作逐渐完善的变化。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A person is painting a watercolor of a cherry blossom tree, with predominantly white and brown tones, in a fresh and natural style. The scene is presented in hand-drawn animation.

步骤 02 生成的视频效果：

这是 Sora 生成的一段画家用笔在画布上作画的视频效果。虽然这不是真实的物理交互，但 Sora 通过模拟物体间的动作和变化，可以在生成



的视频中呈现出类似于物理交互的效果，从而使视频效果更加生动。

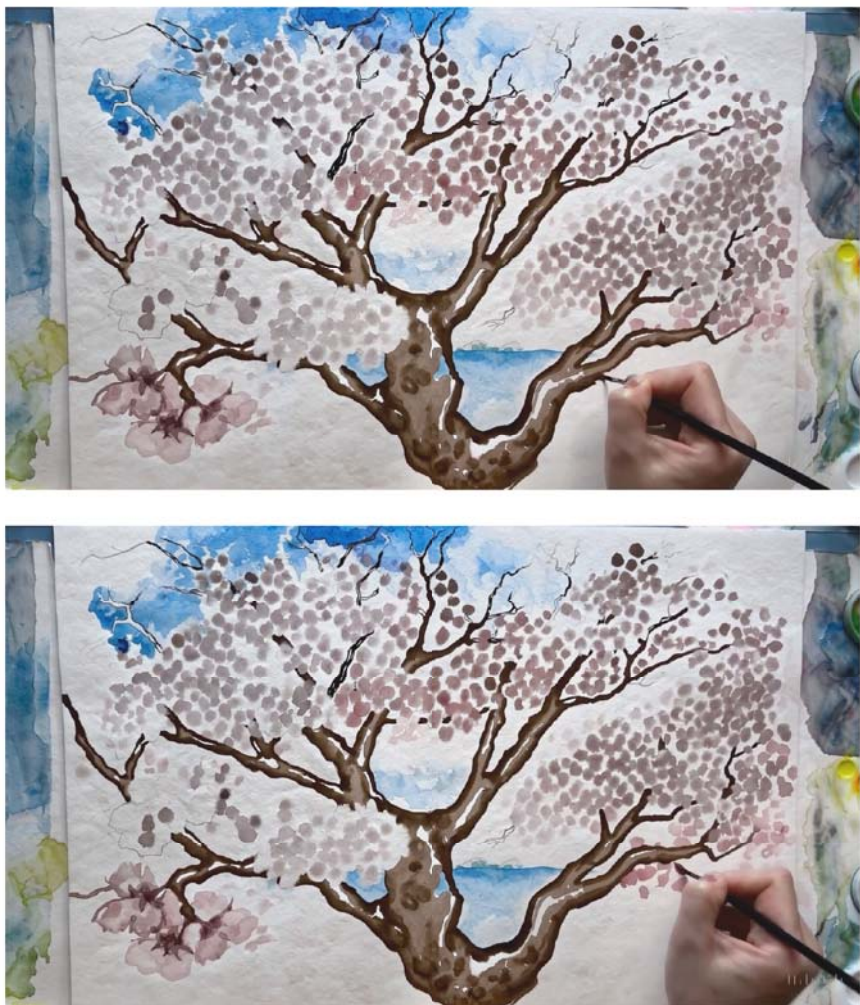


图 2-10 画家用笔在画布上作画视频

中文大意：一个人正在用笔画一幅樱花树的水彩画，色调以白色和棕色为主，采用清新自然的风格，画面以手绘动画方式呈现。

在这个案例中，这些笔触不仅在当时可见，而且会随着时间的推移而持续存在。这意味着画家可以在之前的作品上进行叠加或修改，创造出更加丰富的画面效果。Sora 通过模拟画家在画布上作画的过程，从而在生成的视频中展现出新的笔触，这种效果是通过算法和模型来实现的，而不是通过真实的物理交互。

尽管如此，这种视觉效果仍然可以给观众一种仿佛是通过真实物理交互而产生的感觉，增强了视频的真实感和沉浸感。

2.10 能够模拟数字世界的行为

Sora的模拟能力不仅限于现实世界,它同样可以模拟数字世界中的“人工过程”。Sora模拟的“人工过程”通常指的是Sora能够理解和模拟那些原本需要人类参与或控制的过程。在《我的世界》这款游戏中,Sora不仅可以运用基本策略来控制玩家的行动,还能以极高的保真度呈现出游戏世界及其动态变化,这种双重能力的结合,使Sora在游戏模拟领域具有巨大的潜力。

例如,在视频游戏的情况下,Sora可以模拟玩家在游戏行为,这些行为原本是由人类玩家通过控制器或键盘鼠标等输入设备来执行的。通过学习和模拟这些“人工过程”,Sora可以在没有人类直接参与的情况下,自主地与游戏环境进行交互,并产生类似于人类玩家的游戏行为,如图2-11所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词:

Simulate the game scene of “Minecraft” .

步骤 02 生成的视频效果:

这是 Sora 生成的一段模拟《我的世界》游戏场景的视频效果。



图 2-11 模拟《我的世界》游戏场景

中文大意: 模拟《我的世界》游戏场景。

值得一提的是，Sora 的这些功能可以通过零样本学习的方式来实现。这意味着，在没有任何先验知识的情况下，只需通过简单的提示词，如带有 Minecraft（《我的世界》）的文本指令，Sora 就能够理解并模拟出与该游戏相关的行为和场景。

2.11 打造出逼真的虚拟场景

Sora 展现出了卓越的复杂场景和角色等元素生成能力，它能够轻松生成包含众多角色、各种运动类型、主题鲜明、背景细节丰富的复杂场景。

无论是生动的角色表情，还是复杂的运镜技巧，Sora 都能游刃有余地创造出来，这使其生成的视频不仅具有高度逼真性，还具备引人入胜的叙事效果，让观众仿佛置身于一个真实而又充满故事的世界中，相关示例如图 2-12 所示。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

Two people walking up a steep cliff near a waterfall and a river in the distance with trees on the side, stunning scene.

步骤 02 生成的视频效果：

这是 Sora 生成的一段令人惊叹的山水风光视频效果。



图 2-12 令人惊叹的山水风光视频

中文大意：两个人走上陡峭的悬崖，悬崖边上有一个瀑布，远处有一条河流，旁边有树，这是令人惊叹的一幕。

从图 2-12 中可以看到，Sora 视频采用快速摇镜头的运镜方式，展现出大篇幅的横向场景，而且这些画面能够无缝衔接。Sora 高超的运镜技巧，使画面在快速切换时仍能保持高度的连贯性。这种效果在动作场景、风景描绘或大型活动中尤为显著，能够让观众仿佛身临其境，感受到无与伦比的沉浸感。

2.12 强大的角色动画生成能力

Sora 还具备强大的角色动画生成能力，它能够模拟人物、动物等角色的动作和表情，使生成的视频内容在角色表现上更加生动和有趣，相关示例如图 2-13 所示。无论是角色的动作还是表情，Sora 都能够准确地模拟出这些状态，让角色更加逼真地展现在人们面前。

Sora 的强大能力不仅体现在单个角色的塑造上，更体现在对整个场景的全面掌控上，它甚至还能够模拟不同场景之间的交互行为和联系，使生成的视频内容在场景转换上更加流畅和自然。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A corgi vlogging itself in tropical Maui.

步骤 02 生成的视频效果：

这是 Sora 生成的一段柯基犬在热带毛伊岛拍摄视频的视频效果。



图 2-13 一只柯基犬在热带毛伊岛拍摄视频



图 2-13 一只柯基犬在热带毛伊岛拍摄视频（续）

中文大意：一只柯基犬在热带毛伊岛拍摄视频。

总之，无论是从室内到室外，还是从白天到黑夜，Sora 都能够准确地模拟出这些场景转换，为用户带来连贯而完整的视觉体验；无论是角色的动作、表情，还是场景的布局、光影，Sora 都能精准把握，呈现出令人惊叹的视觉效果。这种对细节的极致追求，使 Sora 在视频生成领域独树一帜。

2.13 呈现出丰富的多镜头效果

Sora 具备在同一样本中生成同一角色的多个镜头的能力，这意味着在整个视频中，同一角色的外观、动作和表情都能得到一致的保持，这一特性对于制作需要多个角度、多个场景呈现同一角色的视频来说尤为重要。

Sora 的多镜头生成能力，在电影预告片、动画以及其他需要多视角展示的场景的制作中尤为实用。通过 Sora，用户可以灵活地在不同镜头间切换，展现角色的不同面貌和动作，同时保持整体视觉效果的连贯性和一致性，相关示例如图 2-14 所示。这种多镜头生成技术不仅提升了视频制作的效率和灵活性，还为观众带来了更加丰富和多样的视觉体验。

▶ Sora 案例生成 ◀

步骤 01 输入的提示词：

A movie trailer featuring the adventures of the 30 year old spaceman wearing a red wool knitted motorcycle helmet, blue sky, salt desert, cinematic style, shot on 35mm film, vivid colors.

步骤 02 生成的视频效果：



这是 Sora 生成的一段电影预告片的视频效果。



图 2-14 一段电影预告片视频

中文大意：一部电影预告片，讲述了一位 30 岁的太空人的冒险故事，他戴着红色羊毛编织的摩托车头盔，在蔚蓝的天空和盐湖沙漠中，采用电影风格，使用 35 毫米胶片拍摄，色彩鲜明。

在电影预告片的制作中，通过 Sora 的多镜头生成能力，能够快速生成多个具有紧张感和悬念的镜头，将观众带入到电影的氛围中；在动画制作领域，Sora 可以轻松创建出多个角度、不同视点的镜头，使动画角色和场景更加生动和立体。Sora 的应用不仅能够缩短视频制作周期，还提高了作品的质量和观赏性。

总体来说，Sora 的多镜头生成能力为视频制作领域带来了革命性的变革，它使

用户能够以前所未有的方式展示角色和故事，为观众呈现出更加精彩、丰富的视觉盛宴。随着技术的不断发展，我们有理由相信，Sora 将在未来的视频制作领域发挥更加重要的作用。

总之，Sora 这些强大的功能展示了视频制作路径的广阔前景：通过不断扩展和提升视频模型的规模和性能，有望开发出能够高度模拟物理世界和数字世界的先进模拟器。这些模拟器将不仅能够精准地再现现实世界中的物体、动物和人物，还能深入模拟它们在各种环境下的行为、互动和演变。

这将为人们提供一个全新的视角和工具，用于研究现实世界的复杂系统，探索未知领域的可能性，以及创造更加丰富和逼真的虚拟体验。随着 AI 技术的不断进步，有理由相信，未来的视频模型将在模拟物理和数字世界的道路上取得更加辉煌的成就。

本章小结

本章主要介绍了 Sora 的 13 个创新能力，包括生成长达 60 秒的超长视频、画面具有丰富的细节、能准确理解自然语言、具有较高的分辨率和灵活的采样能力、保持视频画面与 3D 一致性、保持主体和场景的一致性、能够“与世界进行互动”、拥有强大的角色动画生成能力等。通过对本章内容的学习，读者可以熟练掌握 Sora 的一些创新功能，方便后续更好地应用 Sora 制作出精彩的 AI 短视频。

课后习题

鉴于本章知识的重要性，为了让大家能够更好地掌握所学知识，下面将通过课后习题进行简单的知识回顾和补充。

- (1) 请简述 Sora 生成高质量的视频效果主要依靠哪些技术。
- (2) 请简述 Sora 在生成视频过程中的“长期一致性”特点。



课后习题 1



课后习题 2