

证券研究报告 | 传媒行业 | 2023年9月12日

传媒团队·行业深度报告

3D：AI的下一个涌现

分析师

杨晓峰

登记编号：S1220522040001

杨昊

登记编号：S1220523070004

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

一、文生2D复盘：“千万级数据和亿级数据”是关键

我们认为亿级数据量将是3D出现涌现能力的一个重要突破节点。

在梳理了2D图像发展路径之后，我们发现有三个重要节点值得关注。

- 1) **领跑模型出现**：扩散模型逐渐领先于GAN、自回归模型，成为2D生成模型中主流模型；
- 2) **首个千万级2D数据集出现**：ImageNet数据集已拥有22000类约1500万张标注图像，基于ImageNet的比赛ILSVRC诞生了AlexNet、VGG、GoogleNet等经典的深度学习网络模型；
- 3) **破亿数据集的出现**：LAION数据集通过爬虫突破上亿数据量，Stable Diffusion基于LAION-5B的子集训练得到。

二、3D研究框架：已破千万级数据集，OpenUSD加速数据集扩张

未来，3D的发展将沿着如下思路进行：OpenUSD统一数据表现形式和文件格式，使得3D数据集的量得以大幅提升，而更大数据集和更好模型相互促进发展。

- 1) **首个千万级数据集出现**：7月11日发布的Objaverse-XL数据集包含1020万3D资产，主要来源于网络爬虫，是目前最大的3D数据集。3D生成出现了具有统治地位的NeRF模型和千万级数据集Objaverse-XL，3D发展到达2D图像的“2020-2021年”。
- 2) **数据集持续扩大**：USD简化3D文件访问，统一文件格式，提供了发展加速度。3D模型表示方式多样，其中隐式表示越来越受重视；同时，3D数据文件多样，通用格式解决模型互导问题。

三、文生3D的方向-扩散模型+NeRF

- 1) 目前4类3D生成路径从易到难：扫描得3D、视频得3D、图生3D、文生3D。
- 2) 文生3D是终极状态：文生3D使得3D资产从“专业的”变成“大众的”，是其他三种路径的最终形态。目前已有的算法中，谷歌提出DreamFusion，是较早一批实现了“文生3D模型”的AI生成模式。和DreamFusion相同，其他文生3D模型的主要结构也是“图像训练模型+3D生成模型”的结构。NeRF在3D生成模型领域具有统治地位。
- 3) 近两年，谷歌、英伟达、OpenAI、Meta、Apple公司都有“AI+3D”布局，对于3D的探索路径主要集中在成熟度不高的图片生成3D和文生3D领域。

四、文生3D成本测算：迭代次数万级以上

- 1) 2D的迭代次数：以2D为例，一般情况下，2D场景迭代20-50次左右即可实现要求。
- 2) 3D的迭代次数：单个3D的AI生成目前迭代约30000次左右，选择文生3D模型-Zero123在单个3090显卡上运行，单场景3D资产需要3-4小时左右生成。
- 3) 一个3D资产生成算力成本约5元左右。假设未来迭代速度变快2倍，生成一个3D资产算力花费约2.6元左右，相比于市场3D资产购买费用，有巨大价格优势。

五、投资建议：

3D对VR/AR游戏品类的影响：MR游戏，建议关注宝通科技、恺英网络、恒信东方；

3D对美术密集型品类的影响：开放世界游戏，建议关注网易；动画产业，关注光线传媒、奥飞娱乐；

3D对一般3D游戏公司的影响：3D游戏，建议关注2022年三季度低基数的三七互娱、姚记科技、凯撒文化、星辉娱乐；其他优质游戏公司如吉比特、神州泰岳、盛天网络、巨人网络、游族网络、完美世界。

六、风险提示：

技术发展不及预期，产业推进不及预期，法律及道德风险。

- 一、文生2D复盘：“千万级数据和亿级数据”是关键
- 二、3D研究框架：已破千万级数据集，OpenUSD加速数据集扩张
- 三、文生3D的方向-扩散模型+NeRF
- 四、文生3D成本测算：迭代次数万级以上

一、文生2D复盘：“千万级数据和亿级数据”是关键

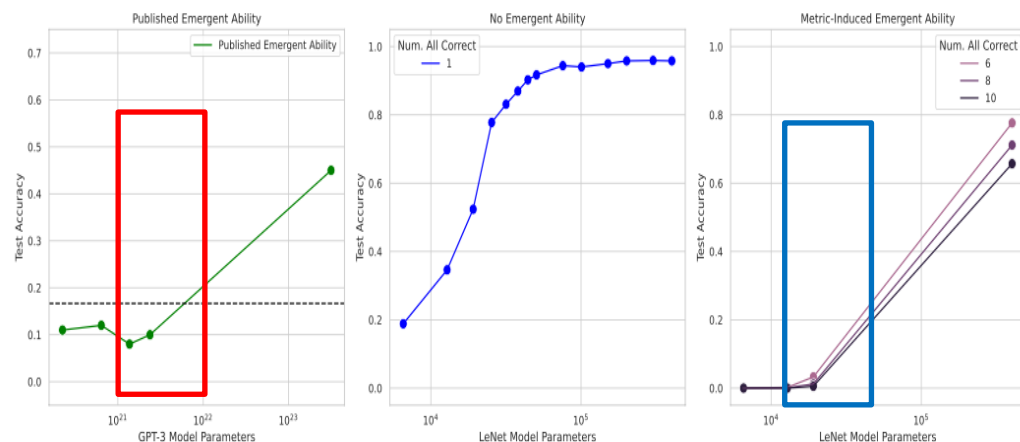


1.1 2D模型数据量变大，涌现能力出现

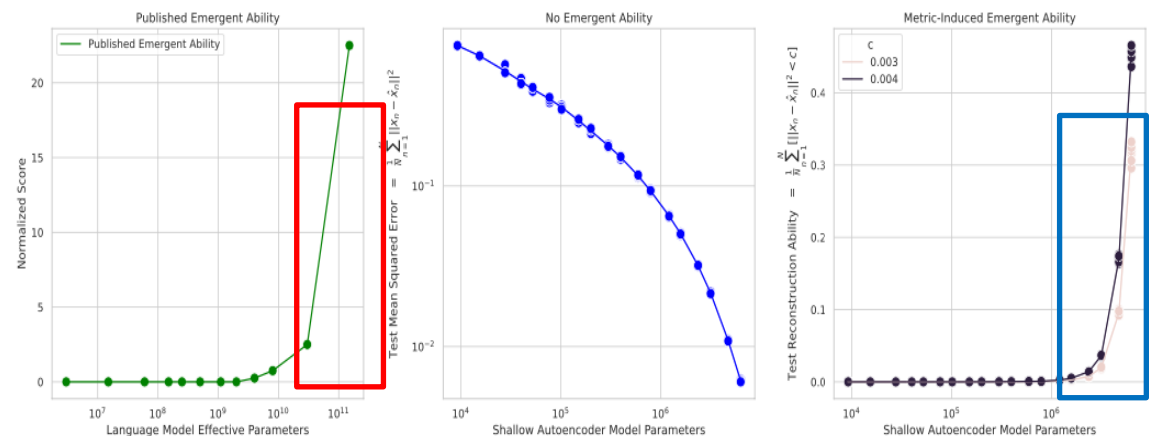
涌现能力的定义：在小模型中没有表现出来，但是在在大模型中变现出来的能力。涌现能力大概可以分为两种：通过提示就可以激发的涌现能力和使用经过特殊设计的prompt激发出的新的能力。无论哪种涌现能力，当模型规模超过一个临界值时，效果会马上提升。

以LeNet和AE为例，涌现能力在CV计算机视觉任务中也有所体现。 LeNet是由Yann Lecun（2018年图灵奖得主创造，其手写体字符识别领域创造性引入卷积神经网络，是CNN的发展起点，通过观察LeNet在MNIST 21上训练的结果，随着模型参数的增加，测试精度显示出不可预测的大幅增长；而在CIFAR100图像数据集上训练的自编码器（AE）随着模型参数的增加显示陡峭的均方重建误差变化。

图表：卷积网络中MNIST分类能力的分类能力涌现
(红色为GPT-3出现的涌现，蓝色为LeNet出现的涌现)

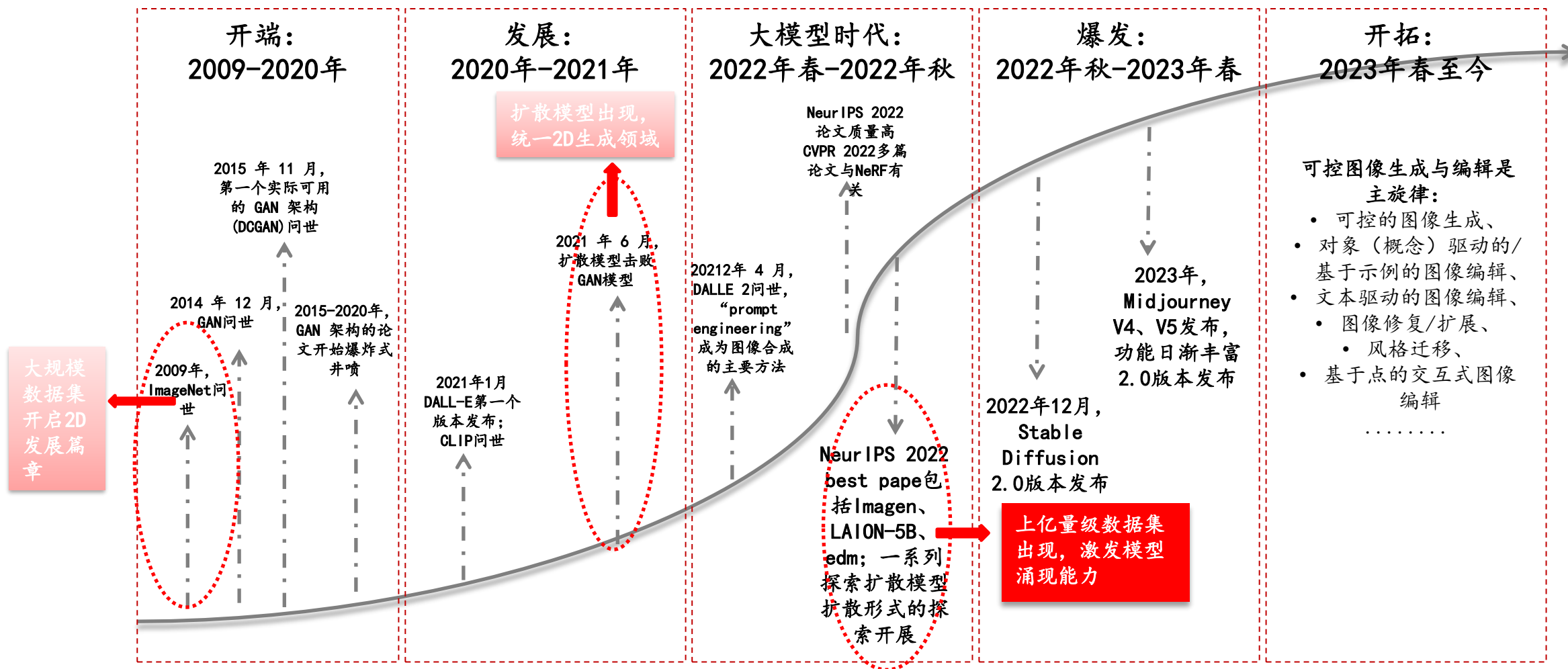


图表：浅层非线性自编码器中的重构能力涌现
(红色为GPT-3出现的涌现，蓝色为浅层非线性自编码器出现的涌现)



1.2 2D发展重要节点：2022年春-2022年秋

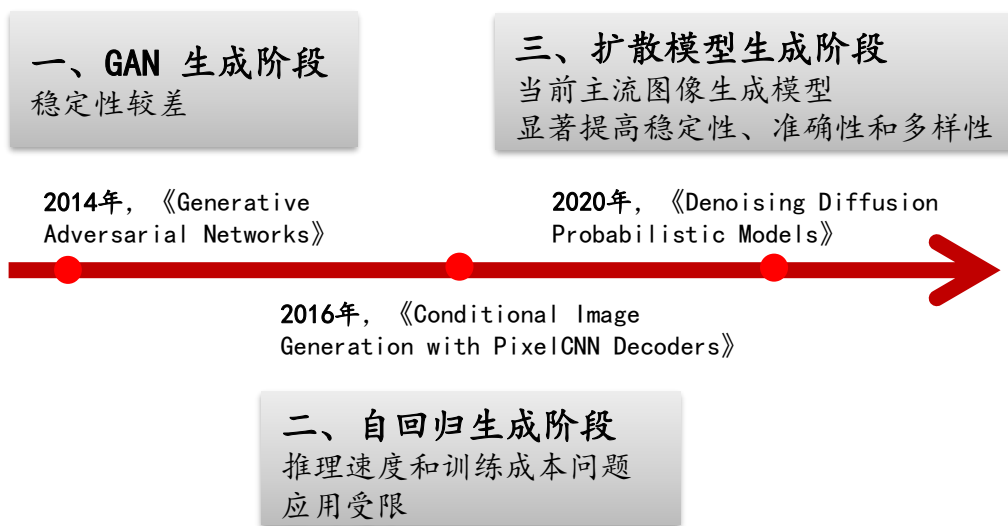
图表：2D图像发展路径



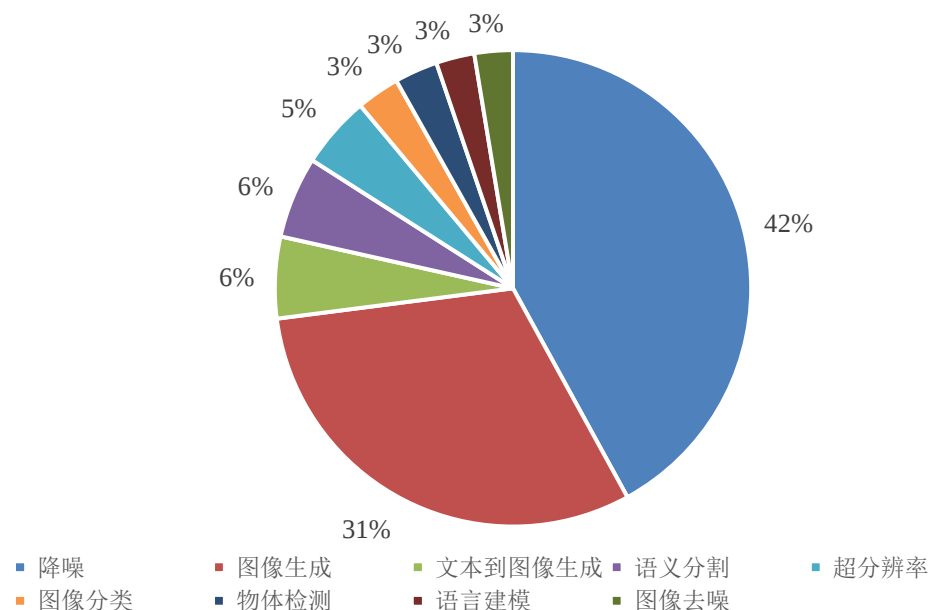
1.2.1 扩散模型领先于GAN、自回归模型

扩散模型（diffusion models）成为2D生成模型中主流模型，逐渐领先于GAN、自回归模型，显著提高了图像生成的稳定性、准确性和多样性。并且扩散模型在诸多应用领域都有出色的表现，如计算机视觉，NLP、波形信号处理、多模态建模、分子图建模、时间序列建模、对抗性净化等。

图表：图像领域主要SOTA模型



图表：扩散模型的主要应用领域



1.2.2 ImageNet：首个千万级别2D数据集，借助众包平台完成

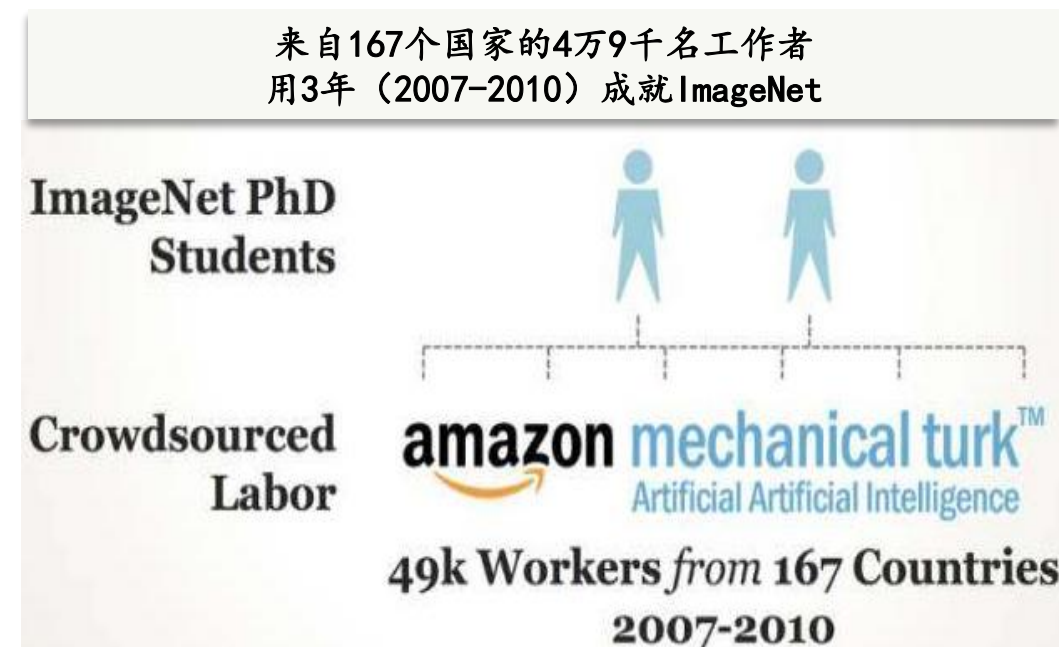
ImageNet数据集09年发布，目前包含约1500万张标注图像。ImageNet数据集由斯坦福大学教授李飞飞牵头从2007年开始收集建立。2009年发布时包含有320万张图像，是当时最大的2D图像数据集。经过发展，ImageNet数据集已拥有22000类约1500万张标注图像。

ImageNet来源于网络图像，借助众包平台（Mechanical Turk）完成。数据集的大量图像数据来自于日益增长的网络平台，考虑到手动标注工作的大量人力需求，研究团队借助于Mechanical Turk（亚马逊开发的众包平台）完成这项工作。

图表：ImageNet数据集出现之前的数据集图片数量

时间	数据集名称	图片数量
1998年	CMU/VASC Faces	337个人超过750,000张图像
1998年	FERET Faces	自1199个不同个体的14,126张图像
1998年	MNIST digits	70,000手写数字图像
1999年	CurRET Textures	5,000+纹理图像
2001年	Middlebury Sterco	几十对立体匹配图像
2003年	CalTech 101	9,146张图像,256个类别
2004年	KTH human action	2,391个人类动作视频
2006年	ESP	1000场景图像
2006年	MSRC	30个人执行12类动作行为
2007年	PASCAL	20类共9963张图片
2007年	Lotus HILL	500,000个图像（或视频帧）
2007年	CalTech 256	30,607张图像,257个类别
2008年	LabelMe	数万张图像及标注
2008年	TinyImages	79,300,000张小型32*32 像素的彩色图像
2009年	ImageNet	300万图像，后包含1500万图像

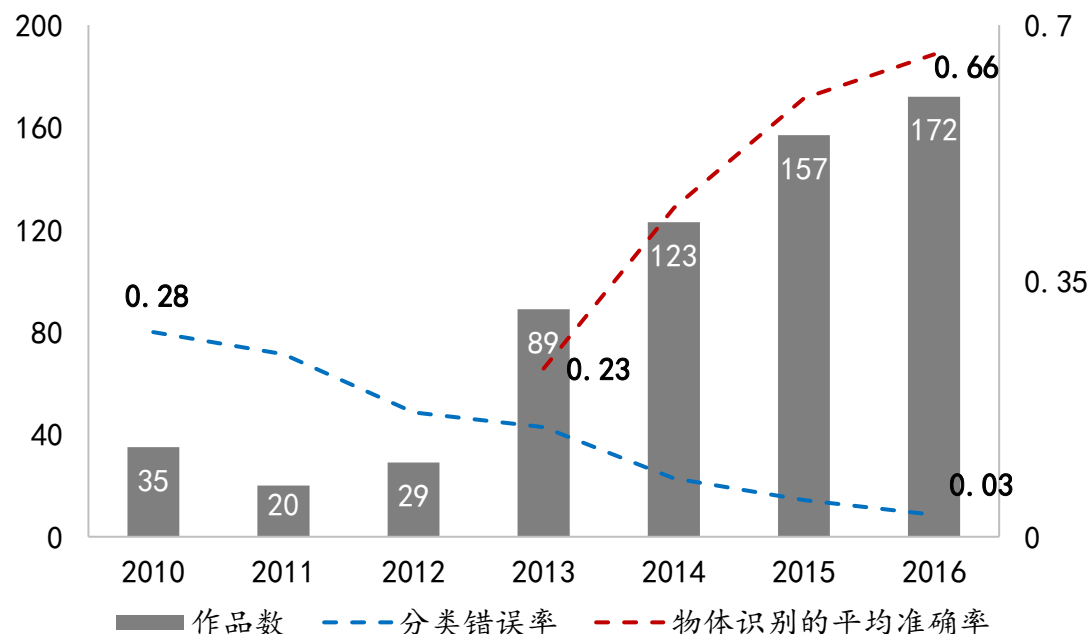
图表：ImageNet借助众包平台完成



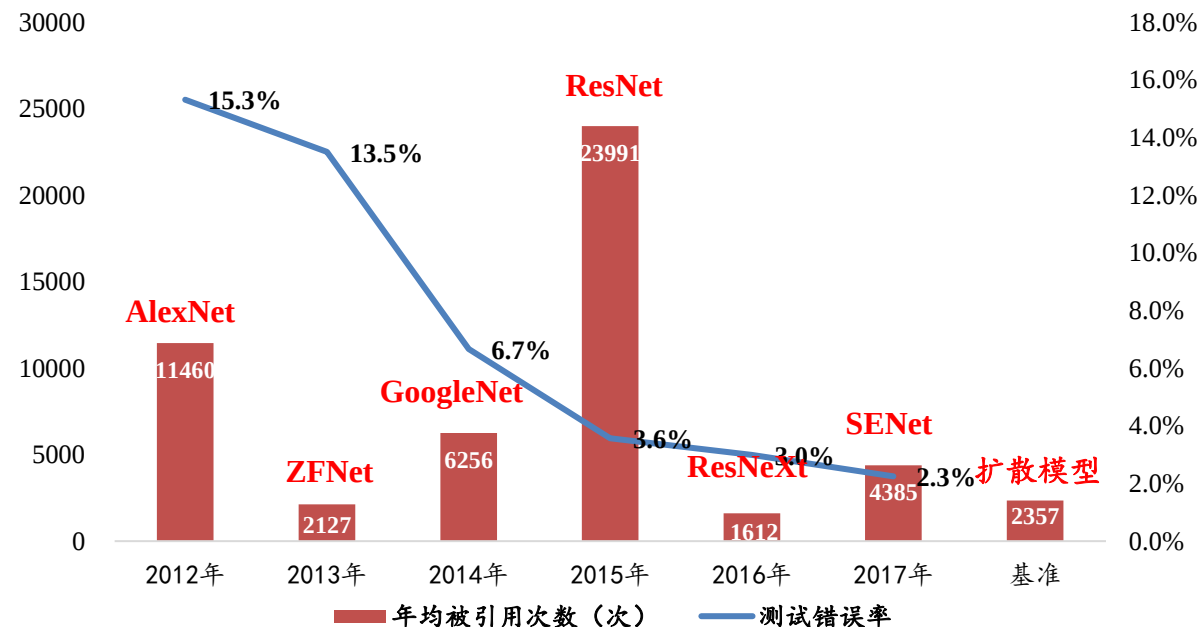
1.2.2 多个重量级算法诞生于ImageNet数据集

基于ImageNet的比赛ILSVRC每年举办一次,从2010年开始举行,2017年后的比赛由Kaggle社区主持。比赛逐渐成为全球大型赛事,2016年有172个作品参赛。短短7年内,分类领域的错误率就从0.28降到了0.03。以扩散模型为基准,ILSVRC冠军算法中AlexNet(12年)、GoogleNet(14年)、ResNet(15年)、SENet(17年)等算法均超过扩散模型的影响力(以年均被引用量表示算法影响力)。比赛的历年优胜者中,诞生了基于ImageNet训练的AlexNet、VGG(2014年定位比赛冠军)、GoogleNet等经典的深度学习网络模型。冠军算法历年的测试错误率在6年间从15%下降到2%,极大地推动了计算机视觉领域的发展。

图表: ILSVRC的参与作品数和TOP 1的分类错误率和物体识别平均值



图表: ILSVRC冠军算法的历年测试错误率和年均被引用量



1.2.3 LAION：破亿数据集出现，模型落地离不开亿级数据

LAION（拥有全球成员的非营利组织）在2021年公布的LAION-400M是当时最大的公开图文数据集,22年10月发布的LAION-5B是上一版本的14倍。网络爬虫带来数亿级别公有数据集。LAION的收集使用加州非营利组织Common Crawl收集到的代码来定位网络上的图片，并将它们与描述性文本关联起来，对人工的依赖性很小。上亿数据集成为图像训练模型成熟应用的必要条件。LAION的数据集已经被用于文生图的生成器，包括谷歌Imagen和Stable Diffusion。同时，其他已落地图像生成应用的国内外公司均是在上亿级别的2D图像数据集进行模型训练取得满意效果。

图表：大规模数据集类型及图片文字对量级

类型	数据集	图片文字对
私有数据集	MS-COCO	330, 000
	CC3M	3, 000, 000
	Visual Genome	5, 400, 000
	WIT	5, 500, 000
	CC12M	12, 000, 000
	RedCaps	12, 000, 000
	LAION-5B	230, 000, 000
公有数据集	CLIP WIT	400, 000, 000
	ALIGN	180, 000, 000
	BASIC	660, 000, 000

图表：国内外图像生成公司&机构模型一览

模型	机构	数据来源
Stable Diffusion	StabilityAI	LAION-5B的一个子集
DALL-E 2	OpenAI	使用CLIP数据和DALL-E数据(共计650M图像)进行训练
Midjourney	Midjourney	开源数据
Imagen	Google	内部数据集（460M图像-文本对），以及Laion数据集（400M图像文本对）

二、3D研究框架：已破千万级数据集，OpenUSD加速数据集扩张

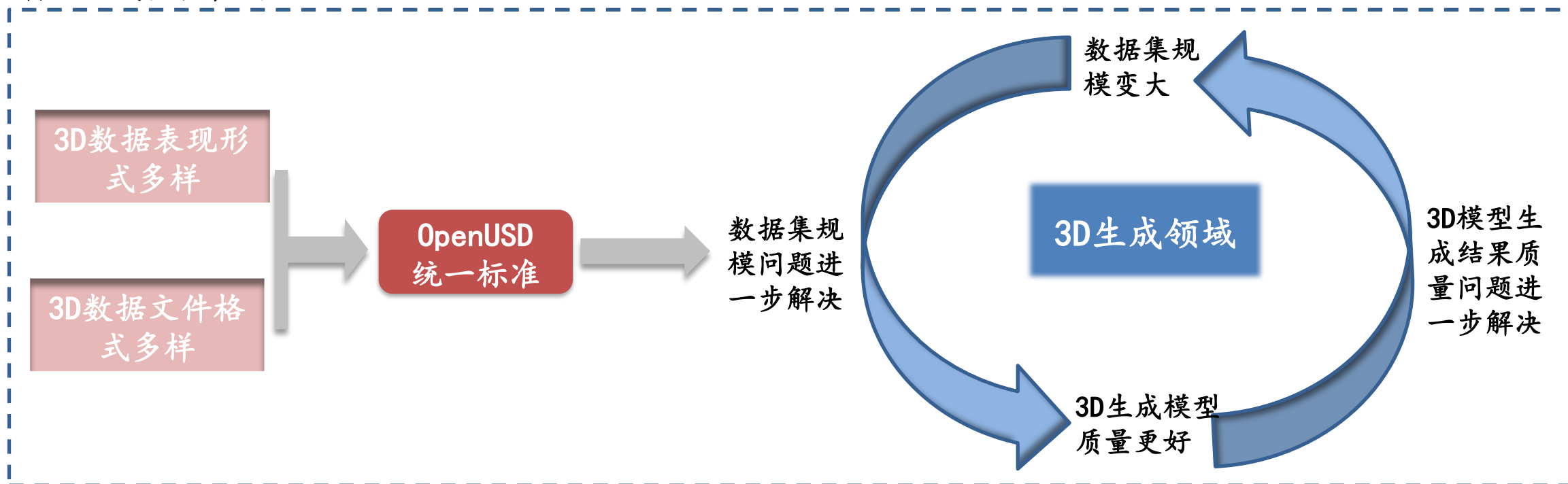


2 3D发展研究框架

OpenUSD统一数据表现形式和文件格式，提供3D发展加速度。3D数据集数据记录形式多样，包含点云、体素等，以往模型往往针对不同形式有不同的算法；同时，3D数据文件格式多样，跨文件转化不易，也提高了数据集收集门槛，限制了3D数据集的发展。而OpenUSD的推进有望解决以上问题，进一步解决数据集规模问题。

更大数据集和更好模型相互促进发展。基于更大3D数据集进行的训练能够得到更好的3D生成质量。生成结果经过简单处理可以大大减少人工建模时间和步骤，从而进一步扩大3D数据集，至此3D生成领域良性循环发展。

图表：3D生成领域发展思路

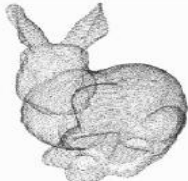
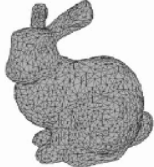
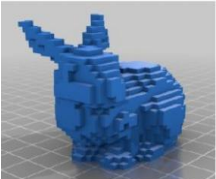
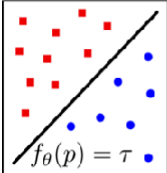
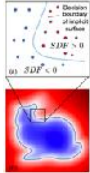


2.1.1 3D模型表示方式多样，隐式表示越来越受重视

主要数据表示方式：隐式表示，其中INRs（隐式神经表示）被用于3D主流模型NeRF中，受到学界重视。3D数据的世界没有一致性，目前已有的3D数据集表示方法包括点云、网络、体素、多视角图片等，不同的表示采用的训练路径也大不相同。

隐式表示解决了显式表示造成重叠、消耗内存等缺点，适用于大分辨率场景，INRs甚至可以生成照片级的虚拟视角。NeRF模型首次利用隐式表示实现了照片级的视角合成效果，也将隐式表示推向新的高度。

图表：3D实例存储的表示方法和收集方式

表示类型	显式表示	显式表示	显式表示	显式表示	隐式表示	隐式表示	隐式表示
模型示例					 	 	
表示方式	点云 Point Cloud	网络 Polygon Mesh	体素 Voxel	多视角图片 Multi-view Images	Occupancy Function	SDF (Signed Distance Function)	INRs隐式神经表示 (Implicit Neural Representations)
表示内容	三维坐标系统中点的集合，这些点通常以x, y, z坐标来表示，并且一般用来表示物体的外表形状。	由一组凸多边形顶点以及凸多边形表面组成的，也叫做非结构化网格。	概念上类似于二维空间中的最小单位——像素	通过不同视角的虚拟摄像机从物体模型中获取到的二维图像的集合	将物体表示为一个占有函数，即空间中每个点是否在表面上	SDF将物体表示为符号距离函数，即空间中每个点距离表面的距离	将信号参数化为一个连续函数，将信号的域映射到该坐标上的属性的值(对于图像，就是将像素坐标映射到R, G, B颜色)
收集方式	激光扫描仪、近红外激光器投影、双目相机拍摄、多视角重建等	使用点云数据转化或将已有3D模型进行转化	不同视角的虚拟摄像机对3D模型摄像或拍摄不同视角图片		使用MLP模拟函数，输入物体3D空间坐标，输出对应的几何信息记录		

2.1.2 通用格式解决模型互导问题，USD简化3D文件访问

3D文件格式多样，适用领域不同，模型互导问题通过通用文件格式解决。3D创建需要各种工具和平台的搭配利用，其他应用程序无法编辑甚至读取，带来了模型互导的问题，导致3D文件格式众多，种类有数十种。多软件之间对同一模型进行编辑操作时先将模型文件另存为通用格式，再导入到其它软件。

USD作为一种通用软件，支持移动端等途径的便携访问和读写。许多传统3D模型为本地客户端设计，需要极高的硬件配置才能正常浏览，存储占用较大且文件打开速度慢。而对于.usd/.usdz文件格式，用户可像浏览普通照片或者文本一样在桌面端、Web端、移动端、AR/VR等终端正常浏览。

图表：常见通用格式及其主要应用方向

特质		.gltf/.gltb	.obj	.fbx	.stl	.3ds	usd/.usdz
包含信息	几何	√	√	√	√	√	√
	材质	√		√		√	√
	动画	√		√		√	√
	图形						
	颜色	√	√	√		√	√
	其他				三角面片		
优点		轻量级、支持复杂的3D数据，包含材质和动画，可嵌入式，可扩展性好	简单易读，广泛支持，适用于交换模型数据	支持动画、材质、骨骼等，可嵌入额外数据	极简格式，适用于机械设计等，数据格式直接，易于转换	广泛支持，能够存储材质和纹理信息	用于特效制作、动画制作、虚拟现实等，支持大规模场景，高性能渲染
缺点		文件可能较大；闭源格式，不适合一些开放场景	不支持动画、不适用于复杂场景，	闭源格式，不适合一些开放场景	不支持材质、纹理、动画等高级信息，不能表达复杂几何结构	不支持现代3D图形特性，不适合高精度渲染	学习曲线较陡，不太适合初学者，某些特性需要高级的计算机图形技能
适用场景		专门为影视和游戏而开发	出现较早，几乎所有知名的3D软件都支持	电影界及视频游戏开发最常用的文件格式	3D打印、工程应用等	专为互联网 Web 而设计	支持跨平台、高性能渲染、可进行移动端访问读写

2.1.3 USD统一3D表示标准, OpenUSD联盟降低USD使用门槛

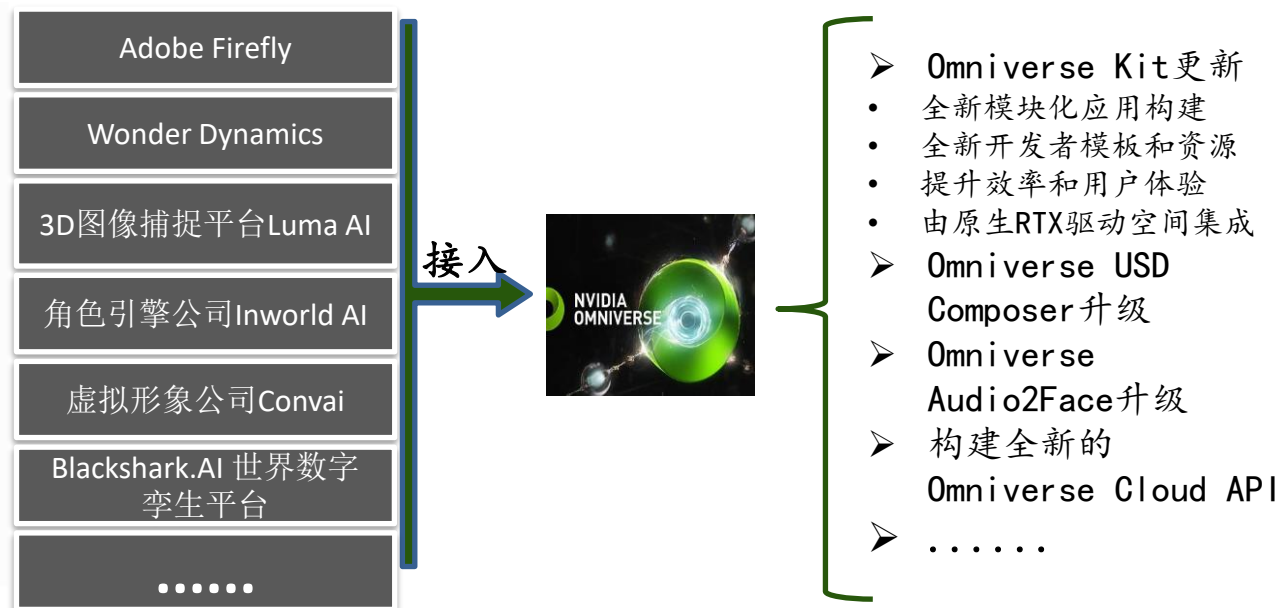
USD承担互联网3D时代的HTML角色。23年8月8日SIGGRAPH大会上, NVIDIA创始人黄仁勋表示:“正如 HTML点燃了2D互联网的重大计算革命, OpenUSD也将开启协作式3D和工业数字化的时代。”

AOUSD不断降低USD文件学习和使用门槛, 推动USD成为3D标准。USD是由皮克斯公司(Pixar)开发的开源格式, 可以在不同的工具间进行内容创作和交换, 但学习曲线陡。针对这一问题, 致力于推动USD文件格式发展的OpenUSD联盟(AOUSD, 由皮克斯、Adobe、苹果、Autodesk和英伟达与Linux基金会下属组成)开展了一系列有效举措, 如NVIDIA Omniverse™ 平台, 来降低用户的使用阻力。

图表: USD支持的3D内容创建应用程序和生态系统名单

• 3Delight	• Revit	• J Cube
• Adobe	• Blender Foundation	• Multiverse for Maya
• Substance 3D Painter	• Blender	• Muse
• Substance 3D Modeler	• Cycles	• Left Angle Autograph
• Substance 3D Stager	• Chaos V-Ray	• Maxon
• AMD ProRender	• E-on	• Cinema 4D
• Apple	• Plant Factory	• Redshift
• Preview and QuickLook	• Vue	• ZBrush
• AR Creation Tools	• Gaffer	• NVIDIA Omniverse
• Rendering Engines	• Golaem	• Procreate
• ModelIO	• Guerilla Render	• Shapr3D
• Motion	• Foundry	• SideFX Houdini
• AnimVR	• Katana	• SynthEyes
• ArcGIS CityEngine	• Mari	• Tilt Brush
• Autodesk	• Modo	• Unity
• 3ds Max	• Nuke	• Unreal Engine
• Arnold	• Intel OSPRay	• Usdtweak
• Fusion 360	• Isotropix Clarisse	• Vicon Shogun
• Maya	• Dreamworks Moonray	• Wizard

图表: NVIDIA Omniverse™ 平台助力OpenUSD 框架和生成式AI

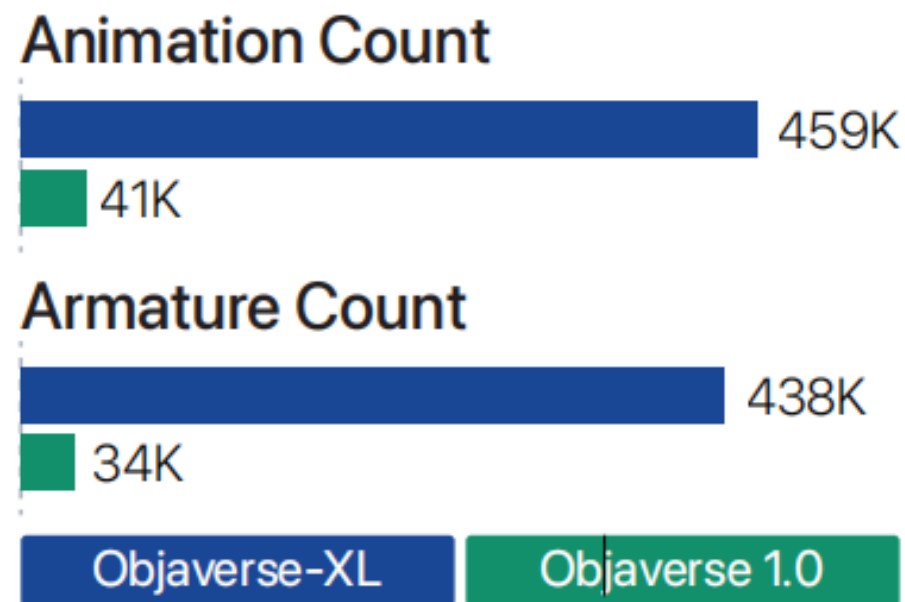


2.2.1 首个千万级数据集出现，3D发展来到2D的“2020-2021年”

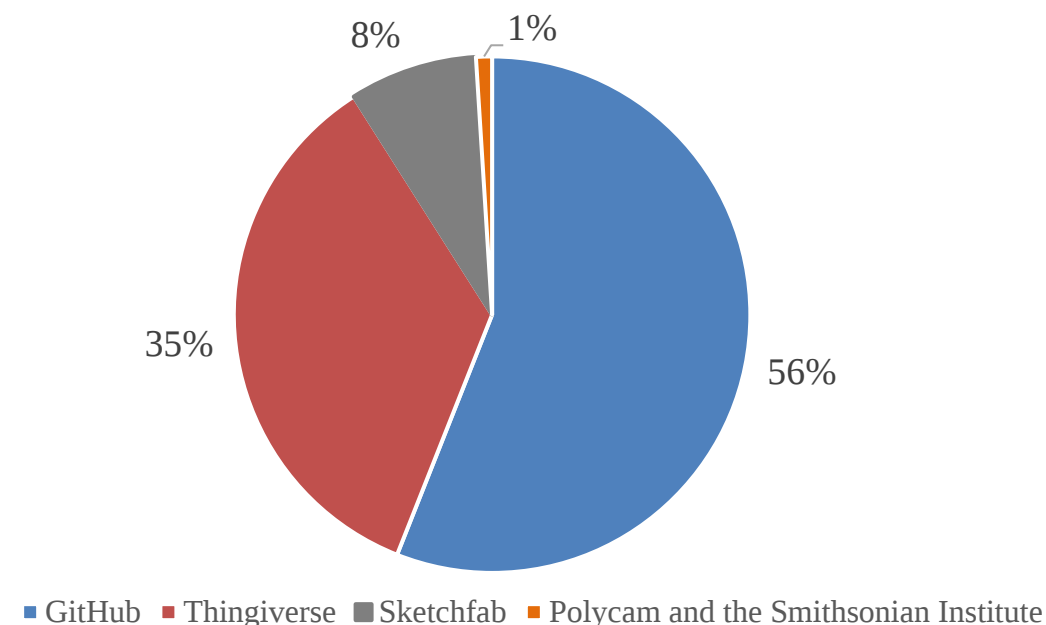
2023年7月11日发布的Objaverse-XL数据集包含1020万3D资产，对比于Objaverse1.0版本增加一个量级。Objaverse-XL通过对互联网上3D对象的类源进行爬虫获取，实例涵盖生活、虚拟多样化场景，最终获得了1020万已渲染的高质量3D文件数据。

类比于2D生成发展路径，3D生成出现了具有统治地位的NeRF模型和千万级数据集Objaverse-XL，3D发展来到2D的“2020-2021年”。可以预见当未来数据集成长到亿级，3D生成也将迎来爆发。

图表：Objaverse-XL比Objaverse大一个量级



图表：3D模型实例爬取来源占比

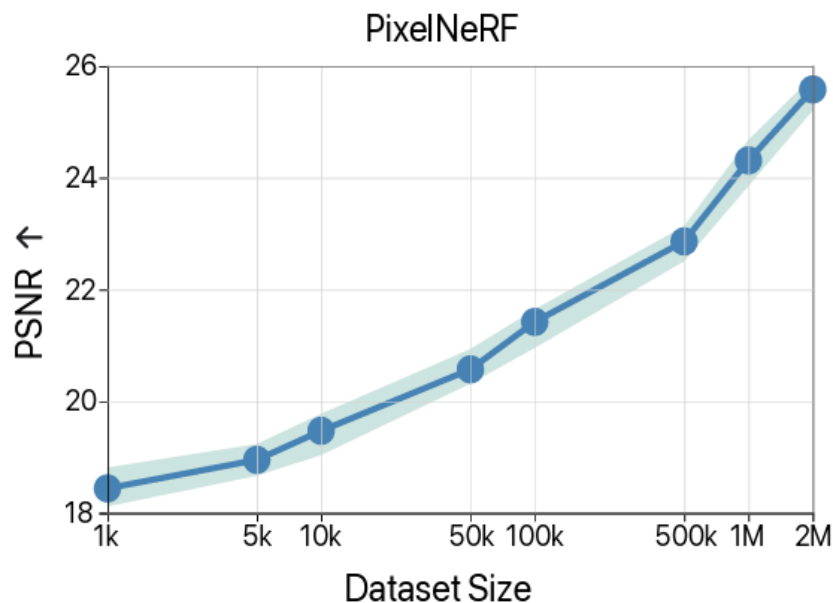


2.2.2 数据集越大，新视角图片更优，3D生成效果更好

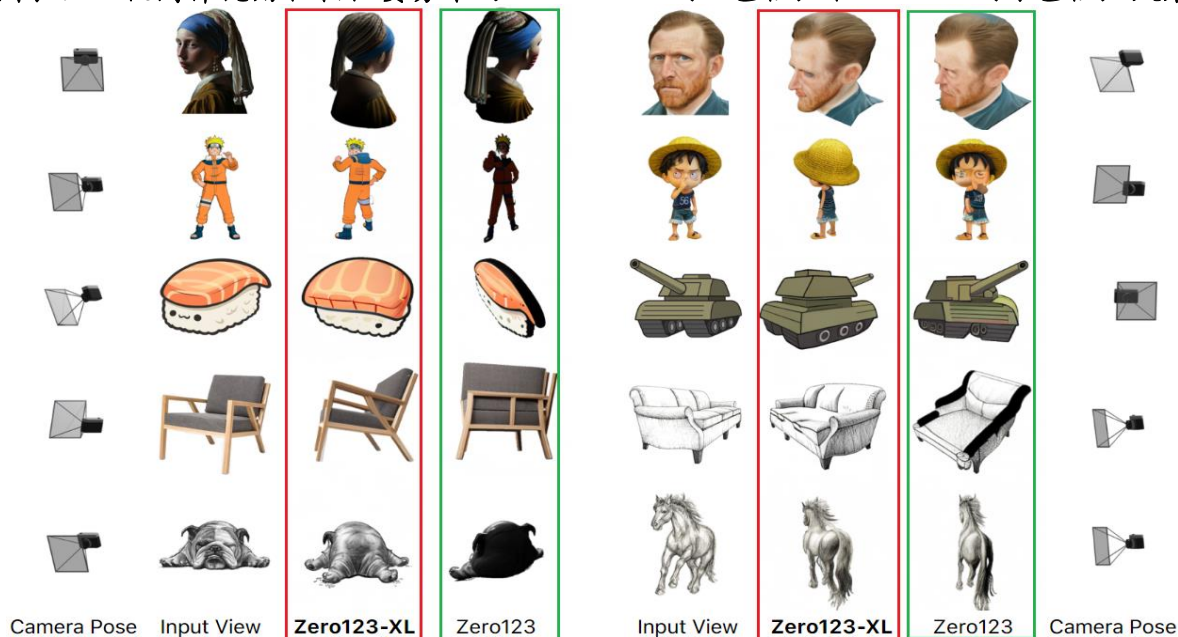
对已有算法PixelNeRF (NeRF的一种改进方法) 和Zero123 (一种文生图模型) 进行训练，发现通过基于更多数据进行训练，3D质量有显著的改进。

PixelNeRF随着数据集量级变大，图像质量评估指标PSNR (Peak Signal-to-Noise Ratio, 峰值信噪比, 数值越大表明失真越少) 变大，表明新生成的视角图像质量越好，有利于提高后续三维重建质量；对使用Objaverse-XL (1000万量级数据)、Objaverse (800万量级数据) 训练得到的Zero123-XL和Zero123比较，发现生成的3D资产侧面、背面的完成度有显著的提升。

图表：PixelNeRF的PSNR与数据集大小关系图



图表：比较同样视角和相机姿势下的Zero123-XL (红色框) 和Zero123 (绿色框) 效果

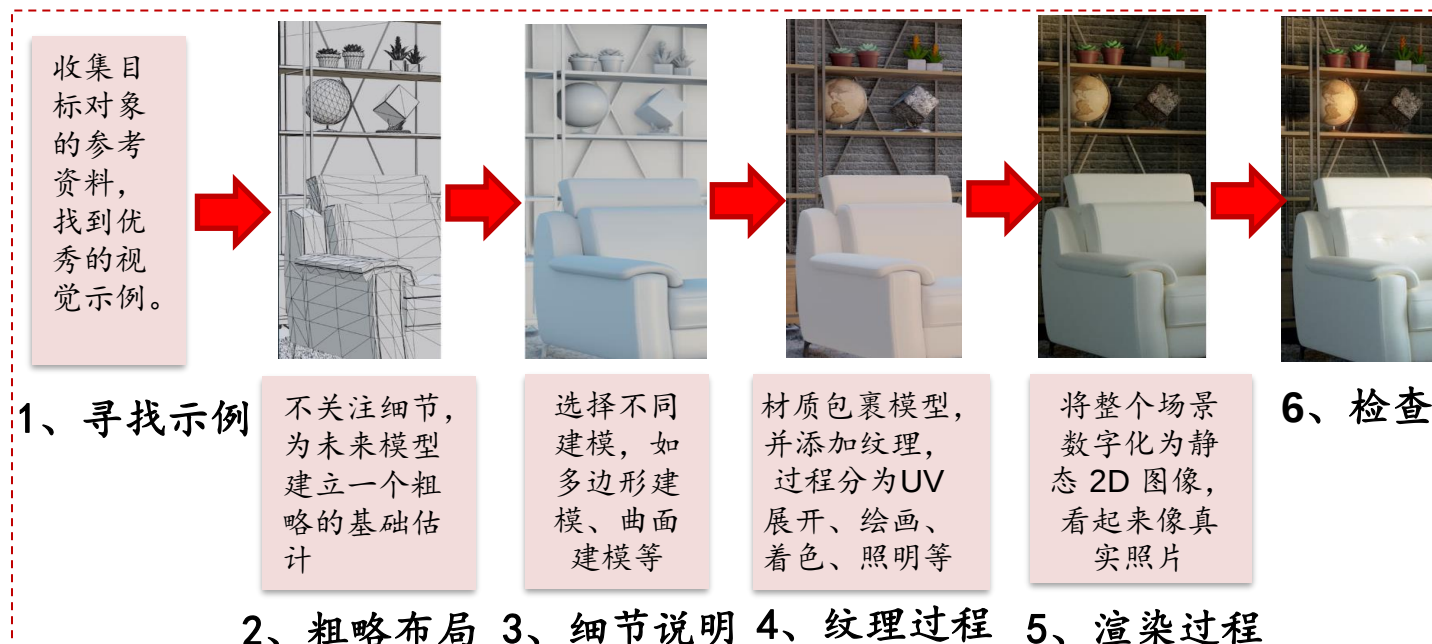


2.2.3 3D资产建模流程长，免费实例不足

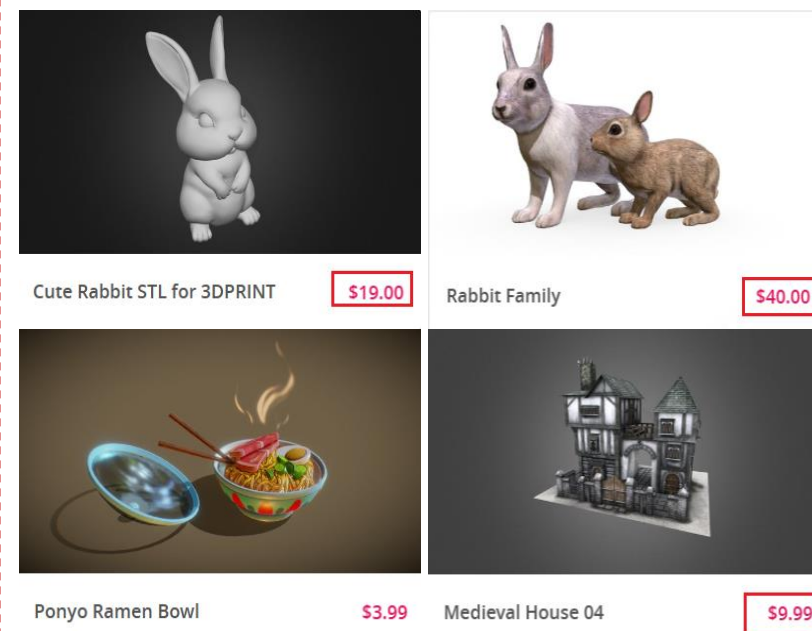
原有3D建模流程长，具有资产属性。3D资产在游戏、动画公司等均属于资产，目前大多数实例由Blender和Maya3D等建模软件手工设计，人工建模涉及环节多，主要流程有：寻找实例-粗略布局-细节说明-纹理-渲染-检查，且建模过程需要大量时间和专业知识。

3D资产昂贵，免费实例不足。公开的免费实例较少，成为阻碍3D数据集扩展的主要原因。在号称世界上最大的3D内容库平台Sketchfab上，3D模型每个价格从\$3-\$500不等。

图表：3D建模流程



图表：Sketchfab部分售卖内容展示

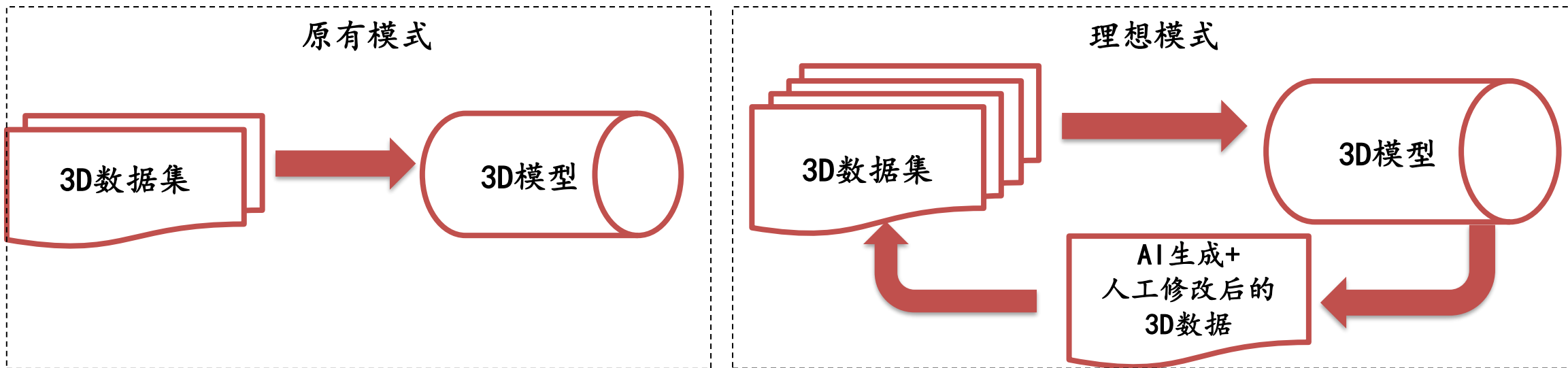


2.2.4 3D生成质量变好，模型产物补充3D数据集内容

3D模型训练效果依赖于3D数据集，但目前的数据集量级无法激发出模型的涌现能力，对生成的AI+3D资产进行人工进一步优化所需要的时间和人力成本依然巨大。

目前数据集仍然是3D生成领域发展的重中之重。更大规模的数据集一旦能够出现，AI+3D模型的质量会随之变好，因此，模型生成的3D资产在经过简单的加工之后即可补充到3D数据集中，良性循环由此开启。

图表：大型3D数据集下训练的高质量3D模型可以有效补充数据集



三、文生3D的方向-扩散模型+NeRF



3.1 3D生成产业方向-文生3D是最终需求

目前存在4类3D生成路径，实现难度从易到难。文生3D是终极需求，将3D资产从“专业的”变成“大众的”。在生成质量、速度、落地程度方面，扫描得3D、视频生3D已经较为成熟，但主要受限于机器的可达性差，没能获得广泛关注；图生3D在特定场景下正逐步落地。相比于前三种方法，文生3D可以直接生成文字描述的场景，甚至现实中不存在的事物，对于没有专业知识的普通人更加友好，且不需要提前准备扫描仪等额外设备。

图表：3D生成路径整理



3.1.1 扫描得3D：成熟落地，手机即可实现

3D扫描技术结合数字三维重建等技术手段，利用三维扫描设备，可以对自然或人工合成物体进行建模，重建真实物体的三维模型，目前已广泛运用于建筑保护、CT扫描、AR/VR等领域。

随着时间的发展，扫描仪的可得性、便携性逐渐增强。2015年安德鲁·塔隆团队就为巴黎圣母院大教堂完成精度5mm的激光扫描和三维建模，越来越多的扫描工作让古物永久保存。2020年iPad Pro出现用于3D扫描和增强现实的深度传感器（LIDAR），配合应用程序3D Scanner Pro可以进行3D扫描。iOS 12的Quicklook功能更是可以用USDZ格式将3D模型分享给其他人。苹果RealityScan依靠手机的摄像头就可以实现扫描建模，iOS端和安卓版本均已发布。

图表：文物三维成像原理



图表：3D Scanner Pro应用场景展示



3.1.2 视频生3D：成熟落地，具备动态实时建模能力

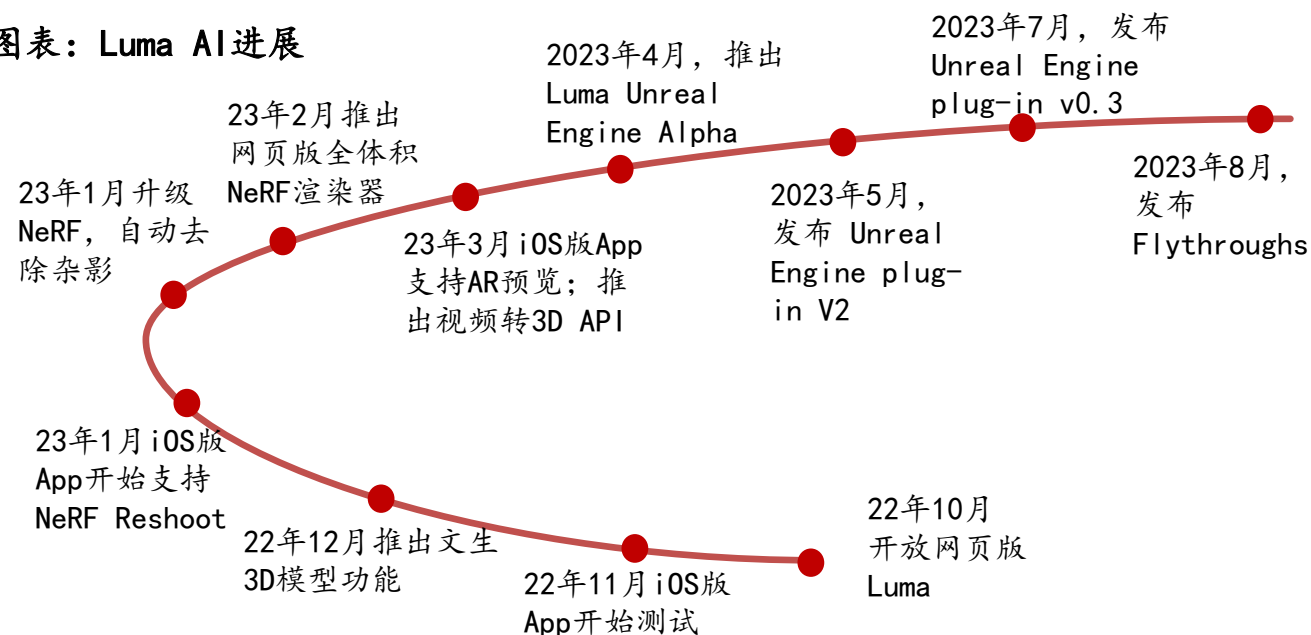
视频生成3D的技术已有落地产品，对于设备的要求越来越低，目前已经可以做到动态实时建模。无需激光雷达或捕捉设备，现在只需要一部手机或者一段视频，即可进行三维场景重建。

22年开放网页版的Luma应用，以NeRF为核心，仅仅通过iPhone即可通过 AI 创建3D；2022年10月，Connect大会上，扎克伯格技术演示中展示了Codec Avatars 2.0的最新进展，Instant Codec Avatars只需用智能手机从不同角度扫描人脸再进行各种表情的扫描，数小时可生成精细的虚拟化身；CYAN.AI平台以CNN（卷积神经网络）和DNN（深度神经网络）为核心，从2D视频中提取人体运动的关键特征进行身体姿势识别，生成的3D成果还可以与Unity集成。

图表：Instant Codec Avatars使用实际生成虚拟化身



图表：Luma AI进展



3.1.3 图生3D：距离真正商用还有一定距离

相比于扫描，图生3D只需要数十张甚至数张真实场景图片即可完成建模。从目前阶段来看，图生3D模型的发展还处于实验室阶段，距离真正商用还有一定的距离。3D生成技术中的NeRF模型的本质就是图生3D。输入是一组二维图像和相应的摄像机参数（包括相机位置和方向），输出是表示三维场景中每个点的颜色和密度的函数。

科技公司在不断尝试研发对应的工具。已有的应用公司包括付费应用Kaedim，用户最少仅需上传一张照片，Kaedim3D即可识别出图中物体的3D形态，并用可导入几乎所有主流软件的格式输出给用户；另外，免费3D模型生成工具PIFuHD可以配合其他2D图像生成软件生成3D模型，但模型并不能直接导入虚幻引擎或者研发引擎当作游戏角色或者NPC。

图表：PIFuHD生成效果：在有些角度质量仍有暇

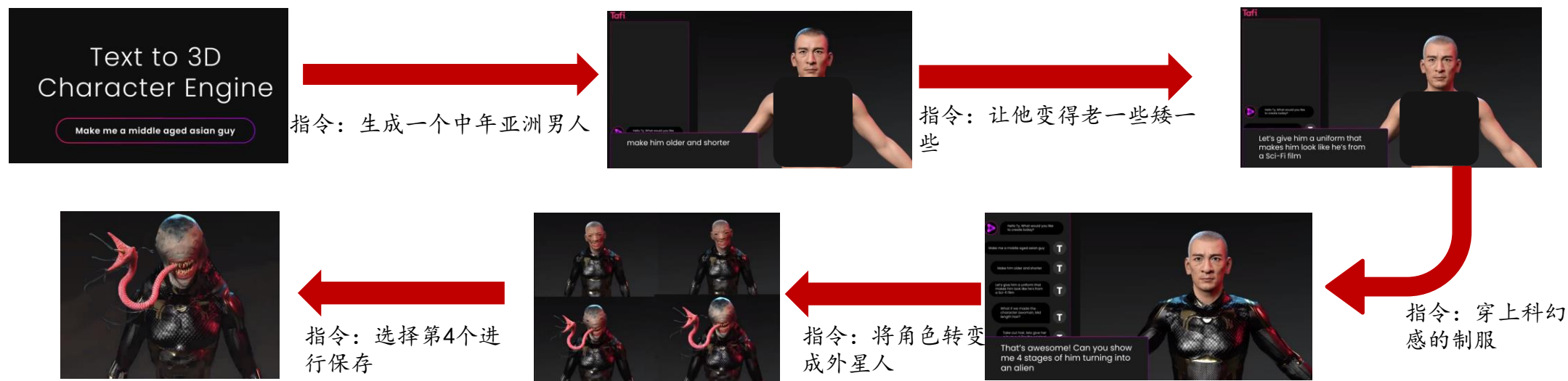


3.1.4 文生3D：学界以科技巨头紧密发力领域

应用端能够支持文生3D的平台很少（如3dfy.ai、Tafi、Masterpiece Studio平台），具有时间长，复杂模型精度低等问题，仍未达到运用于产业生产环节的标准。

2023年6月，Tafi发布文生3D引擎，受限于数据集，其生成内容主要是3D人型角色。领先的3D内容和软件提供商Tafi利用专业美术师支持的原创角色平台“Genesis”庞大的3D数据集，宣布可以从文本创建3D角色，可生成数十亿个3D角色变体。生成的结果可导出到各种 DCC 工具，将高质量角色输出到流行的游戏引擎和 3D 软件应用程序，例如 Unreal、Unity、Blender、Maya、Maxon Cinema 4D等，但受限于训练数据来自3D角色资产，Tafi的生成内容主要集中于人型角色。

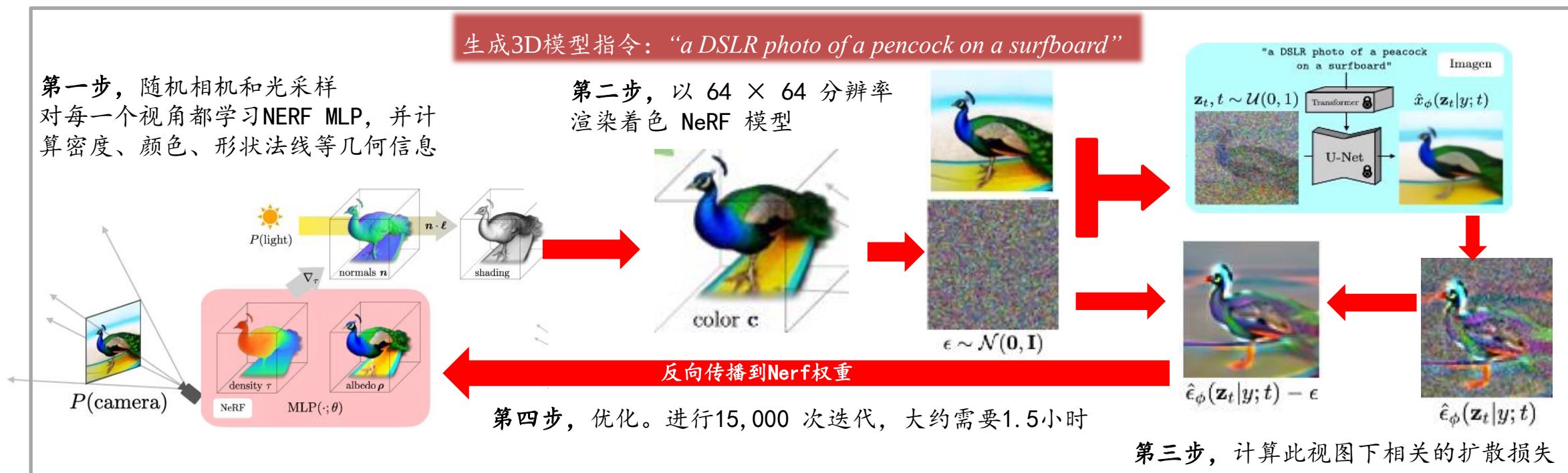
图表：Tafi演示视频



3.2.1 谷歌DreamFusion算法-Imagen+NeRF

DreamFusion由谷歌提出，通过将2D生成模型与NeRF结合实现，是较早一批实现了“文生3D模型”的AI生成模式。底层原理是利用生成2D的图像模型获得Prompt的随机视角视图，在计算误差损失之后优化神经辐射场(NeRF)生成的3D模型，通过对这一过程的循环迭代以达到较精确效果。DreamFusion将大型AI图像模型Imagen与NeRF的3D功能相结合，首次在无需3D数据的情况下完成开放域的文本到3D的合成。

图表：DreamDiffusion的训练过程



3.2.2 “图像训练模型+3D生成模型”是文生3D的主要结构

对于文生3D模型来说，模型一般都具有：1、生成不同视角图片的训练模型；2、利用新视角或者图片进行渲染得到的3D生成模型。其中新视角的生成来达到优化的目的。和DreamFusion同一思路的还有Point-E模型（OpenAI团队）、Magic3D模型（英伟达）、ProlificDreamer模型（清华大学团队）等模型。各视角下，一致性更高的2D生成往往决定了3D资产最初阶段生成的粗模精度；而3D粗模的质量越高，渲染得到精模所花费的时间也会减少。

图表：主要文生3D模型的两部分结构所用模型

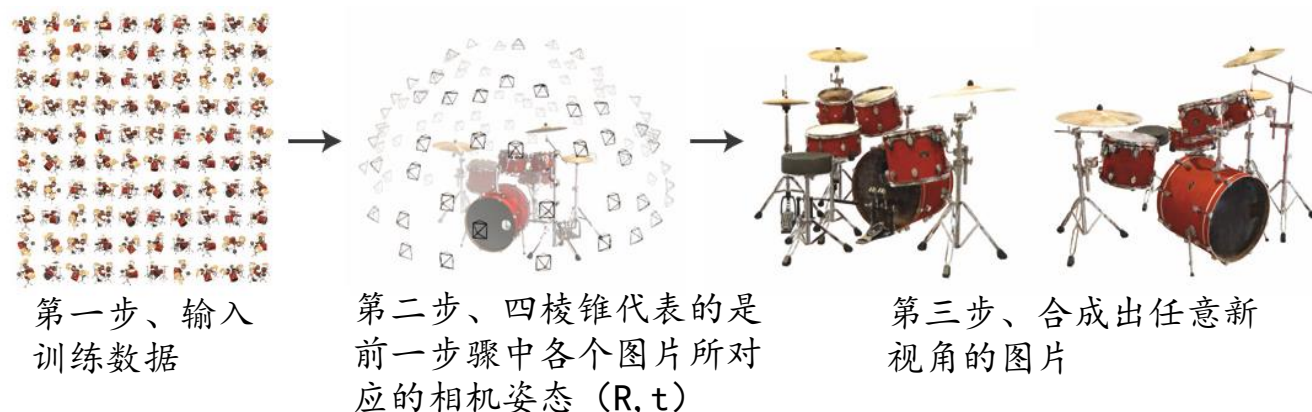
时间	模型	组织	图像训练模型	3D生成模型
2022年11月	Magic3D	英伟达	高分辨率潜在扩散模型（LDM）	NeRF改进模型-Instant NGP
2022年12月	Point E	OpenAI	扩散模型DALL-E2等	基于数百万3D对象和相关元数据训练得到的3D点云扩散模型
2023年5月	ProlificDreamer	清华大学	2D扩散模型如 Stable-Diffusion+基于该预训练模型的LoRA	使用VSD和DMTet进行优化的Instant NGP

3.2.3 NeRF在3D建模领域具有统治地位

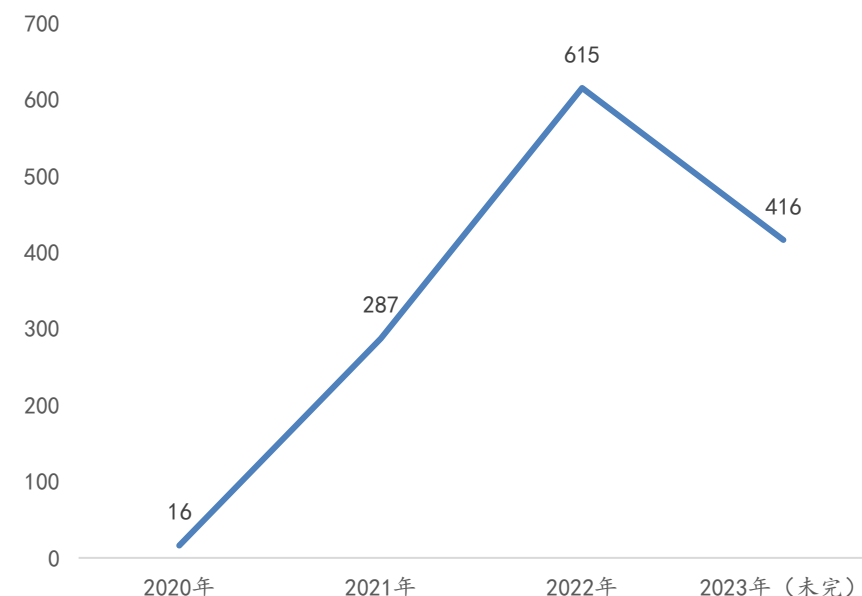
NeRF的研究目的是合成同一场景不同视角下的图像。根据给定一个场景的若干张图片，推理时输入不同视角就可以合成（渲染）该视角下图像。神经隐式表示（neural implicit representation）从2019年开始兴起，2020年ECCV会议上的最佳论文《NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis》提出NeRF算法，将隐式表示推向新的高度。

3D生成领域内，NeRF能力受到广泛认可，具有统治地位。2022年，关于NeRF的引用文章就有615篇，其中不乏高质量文章，在2020-2022年里，NeRF的引用量多达1074次。

图表：NeRF的训练过程



图表：NeRF论文被引用量（次）



3.3 大厂的3D生成探索

近两年，谷歌、英伟达、OpenAI、Meta公司都有“AI+3D”布局，对于3D的探索路径主要集中在成熟度不高的图片生成3D和文生3D领域。23年初，ChatGPT文本生成、2D图像生成的爆发激发了市场对于这两种路径的探索，希望打通利用AI直接从文字生成3D的通道：1、文生2D后，2D生3D；2、文生3D。

- 1、英伟达：人工智能计算公司，在3D领域的布局较早。在22年就有了Get3D、Magic3D等涉及文生3D的模型，是3D领域的重要开拓者。同时不断升级芯片适用于图形计算和3D渲染等，参与到AUSD致力于3D的统一和发展。
- 2、谷歌：作为全球最大的搜索引擎公司，同时也是科技公司。DreamFusion模型是最早一批实现了“文生3D”的模型，为后续公司的探索打下了基础。
- 3、Meta：前身Facebook，在3D领域的探索和打造元宇宙的公司目标相吻合。研究方面，Meta致力于3D虚拟人的突破以及如何搭配VR/AR眼镜的视频转3D技术。
- 4、OpenAI：在23年初的AI爆发中拥有GPT语言大模型和DALL-E文生图模型，并不断探索多模态。3D方面，延续了“文生XX”的模式，充分利用语言大模型优势，贡献有Shap-E、Point-E等文生3D模型。
- 5、Apple：苹果致力于打造“3D-AR”生态的软硬件体系化，其中头显研发已经有七年，且公司不断进行3D领域的技术储备、相关公司收购等，每年的WWDC大会也都会有AR/MR软硬件的消息更新。

3.3.1 大厂的3D生成探索-英伟达：挖掘芯片的3D应用领域

英伟达

2022年9月英伟达发布GET3D模型

3D内容编辑、文生3D

具备能力：能快速生成虚拟世界的对象，可生成具有高保真纹理和复杂几何细节的3D形状，还可以做到文本引导形状生成。

模型概述：过程分为几何分支和纹理分支两部分，前者可输出任意拓扑的表面网格，后者产生纹理场

应用场景：将生成模型导入到游戏引擎、3D建模器和电影渲染器中进行编辑；使用英伟达的另一个AI工具StyleGAN-NADA，可以使用文本提示为3D图像添加特定的风格

图表：生成模型具有相同纹理，不同形状



2022年11月发布Magic3D 模型

文生3D

具备能力：输入文字生成3D；基于提示的 3D 网格编辑

模型概述：两阶段，低分辨率优化（3D重建模型Instant NGP）和高分辨率文本到图像潜（latent）扩散先验（Instant NGP）。

应用场景：产生的3D模型可以应用于虚拟现实、数字双胞胎或者机器人技术等领域。

图表：Magic3D的文生3D功能



3.3.1 大厂的3D生成探索-英伟达：挖掘芯片的3D应用领域

英伟达

2023年8月发布Neuralangelo模型源代码

视频转3D

具备能力：将任何设备拍摄的视频转化为高度详细的3D结构，实现对现实世界中建筑、雕塑以及其他物体或空间的完全复制，保留细节的同时实现更为精确的3D结构恢复；

模型概述：将多分辨率3D哈希网格的表示能力与神经表面渲染相结合；

应用场景：产生的3D模型可以应用于虚拟现实、数字双胞胎或者机器人技术等领域。

图表：Neuralangelo 生成复杂细节和纹理的3D结构



英伟达在SIGGRAPH大会上率先展示了基于AI的3D技术

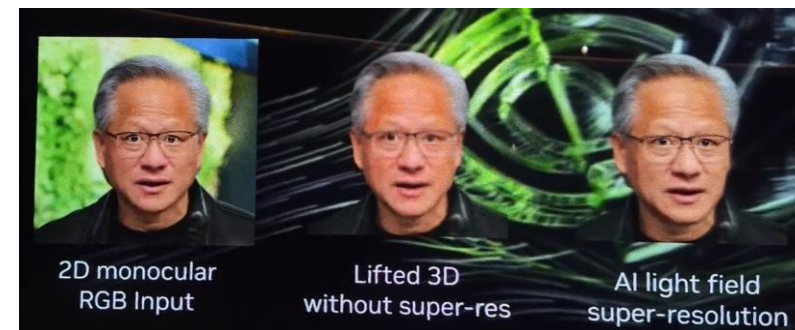
视频转3D

具备能力：AI 3D自拍、实时3D头像，换各种卡通脸，使用头部跟踪显示；3D视频会议

模型概述：关键技术包括3D提升、NeRFs等；

应用场景：主要通过AI技术降低成本，使用普通摄像头就可以实现3D重建和渲染，提供高保真的3D交互体验。

图表：SIGGRAPH 2023大会英伟达演示



3.3.2 大厂的3D生成探索-谷歌：首批文生3D的实现者

谷歌

文生3D

2022年10月发布DreamFusion模型

具备能力：可以在任意角度、任意光照条件、任意三维环境中基于文本提示生成模型，整个过程既不需要3D训练数据，也无需修改图像扩散模型

模型概述：使用预训练2D扩散模型生成二维图像，通过梯度下降法优化神经辐射场NeRF模型

应用场景：产生的3D模型可以应用于虚拟现实、数字双胞胎或者机器人技术等领域。

图生3D

2022年9月谷歌“LOLNeRF”研究结果

具备能力：从 2D 图像中的多个点估计 3D 空间中每个点的颜色和密度，从图像生成 3D 模型的“LOLNeRF”具有更高的精度。

模型概述：使用预测的二维landmarks将数据集中的图像大致对齐到典型的姿态，对于生成模型部分，采用自解码器框架

应用场景：可以通过从汽车、人类等对象集合中提取特征来学习 3D 结构脸和猫。每个感兴趣的对象只需要一张图像，不需要对同一个对象进行两次训练

图表：从一张图来生成3D模型



3.3.3 大厂的3D生成探索-Meta：推进元宇宙的实现

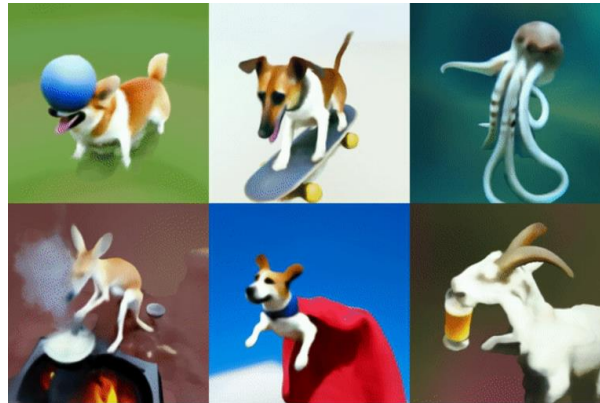
Meta

2023年3月，Meta提出Make-A-Video3D

文生动态3D

具备能力：完成由文本或者图像到 4D（3D+时间）应用的转换
模型概述：运用 4D 动态神经辐射场（NeRF），通过查询基于文本到视频（T2V）扩散的模型，优化一致性。
应用场景：是第一个基于文本描述生成 3D 动态场景的方法，可以为视频游戏、视觉效果或 AR/VR 生成动画 3D 资产

图表：MAV3D直接实现文字生4D



2022年10月展示全身编码器化身Codec Avatars 2.0

扫描得3D

具备能力：只需用智能手机从不同角度扫描人脸再进行各种表情的扫描，几个小时就能生成精细的虚拟化身
模型概述：基于已经训练过的Prior model 根据新的输入图像和深度数据来塑造一个定制化数字人
应用场景：数字虚拟人制作，多种模型的能力不断提高，为Meta元宇宙奠定基础。

图表：iPhone 扫描实现VR化身



3.3.4 大厂的3D生成探索-OpenAI：丰富“文生XX”模式

OpenAI

2023年5月，OpenAI发布Shap-E模型

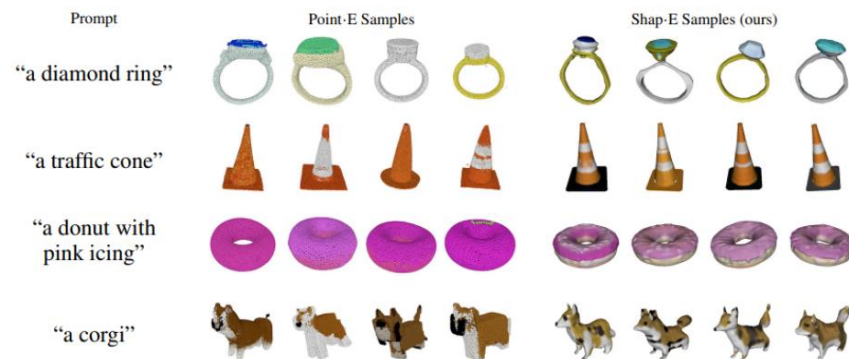
文生3D

具备能力：相比基于点云的显式生成模型Point-E，收敛速度更快，在更高维的表示输出空间中实现了更好的样本质量

模型概述：训练编码器来生成隐式表征+在编码器产生的潜表征上训练扩散模型

应用场景：生成同时表示 NeRF 和网格的 INR，允许它们以多种方式渲染或导入下游 3D 应用

图表：Shape-E和Point-E生成效果对比



2022年12月OpenAI发布Point-E

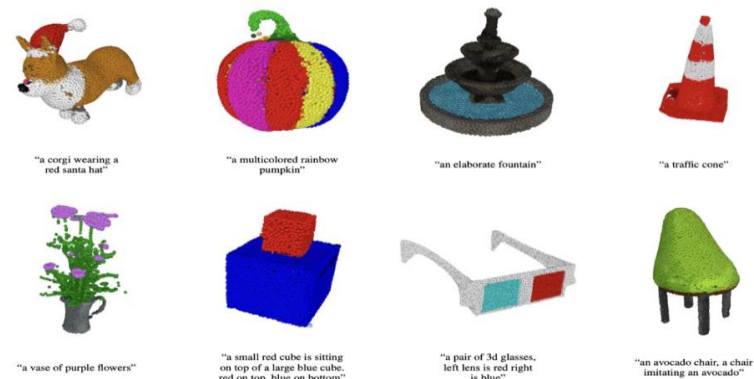
文生3D

具备能力：产生3D点云，或代表三维形状的离散数据点集合

模型概述：文本对图像模型+图像对三维模型。图像对三维模型被输入一组与三维对象配对的图像，在两者之间进行转换。

应用场景：Point-E的点云可以用来通过3D打印制造现实世界的物体。有了额外的网格转换模型，并且更光滑以后，系统也可以找到进入游戏和动画开发工作流的切入点

图表：Point-E点云生成结果



3.3.5 大厂的3D生成探索-Apple: 打造“3D-AR”生态的软硬件体系化

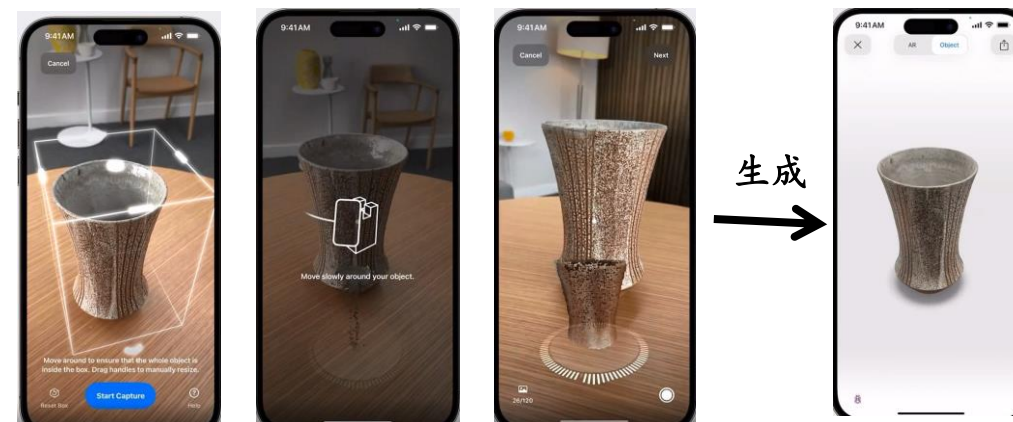
Apple

扫描得3D

图表：Object Capture在iOS上的演示

具备能力：在3D-AR的软硬件生态领域，布局多年。软件方面，推出文件格式USDZ，渲染平台RealityKit、3D文件转换器Reality Converter等；硬件方面，搭载Lidar摄像头、即将推出的Vision Pro空间计算设备等。

应用场景：用更加简便地方式创造和编辑现实转化的3D资产，将3D与日常应用相结合，适用于AR/MR的构建。



发布ARKit框架
帮助开发者在Apple生态下进行AR应用的开发，提供可以识别平面以及跟踪设备移动情况的能力

- 为AR创作提供“USDZ”的文件格式
- 在原有QuickLook的基础上推出AR QuickLook
- ARKit得到增强

- 发布专门用于AR场景的渲染引擎Reality Kit
- 配套的3D内容生产工具：RealityConverter和RealityCompos
- ARKit更新

- Lidar摄像头在一些设备上增加
- 推出Reality Converter
- 新版Apple Map推出3D街景

- iOS15为苹果地图带来全新AR导航功能
- RealityKit2发布
- 推出了Object Capture计划

- 推出Room Plan, 可创建房间的3D平面图

- 发布首个空间计算设备Vision Pro, 搭载vision OS系统, 使用完全3D的界面。
- iOS 17也将支持Object Capture

WWDC 2017大会

WWDC 2018大会

WWDC 2019大会

WWDC 2020大会

WWDC 2021大会

WWDC 2022大会

WWDC 2023大会

四、文生3D成本测算：迭代次数万级以上



4.1 2D生成需迭代约20-50次

以2D为例，迭代次数越多，生成质量越高，对象越复杂。一般情况下，2D场景迭代20-50次左右即可实现要求。根据我们实机测试，利用RTX 3090显卡，在Stable Diffusion中，同一指令来生成的3组2D图片发现：2D模型的迭代速度约为每秒2.1次-4.1次迭代，且速度边际提升。

图表:同一指令不同迭代数和迭代时间



5步迭代，花费2.41秒，速度：2.1次迭代/s



20步迭代，花费5.7秒，速度：3.5次迭代/s



50步迭代，花费12.2秒，速度：4.1次迭代/s

4.2 3D资产算力-生成时间约3-4小时/个

采用3D模型Zero123，单场景3D资产生成需要迭代30000次左右，在RTX 3090显卡（24GB）上运行约3.3-4.2小时。根据我们实机测试的结果，在Zero123算法下，设定迭代30000次时生成的3D成品相对可行。文生图模型的过程首先经过文生图片过程（时间几乎可以不计），再通过生成的新视角图片利用NeRF模型生成3D资产。利用3090显卡（24GB）且显存使用率约50%（即12GB左右）时运行文生3D模型，对处理过的单场景图片，根据测算，每秒迭代次数约2.35次。总优化时间（小时）=迭代次数/每秒迭代次数/60seconds/60minutes。对于需要迭代30000次的单场景来说，生成一个3D模型需要3.3-4.2小时。

图表:不同迭代数和迭代速度下算力时间花费

每秒迭代数 (次)	单场景不同迭代次数下所需GPU（3090）小时数			
	25000	30000	35000	40000
1	6.9	8.3	9.7	11.1
1.5	4.6	5.6	6.5	7.4
2	3.5	4.2	4.9	5.6
2.5	2.8	3.3	3.9	4.4
3	2.3	2.8	3.2	3.7
3.5	2.0	2.4	2.8	3.2
4	1.7	2.1	2.4	2.8

4.3 3D资产算力-生成成本约5元/个

使用RTX 3090显卡在Zero123模型下迭代30000次，生成一个3D资产的算力成本约5元左右。若在RTX 3090上运行文生图3D模型Zero123，按照RTX 3090显卡每小时1.39元租金计算，若每秒迭代2.25次，需要迭代30000次，则3D资产算力成本=5.15元（30000次/2.25次/s/60min/60sec*1.39元/hour）。未来随着模型的发展成熟，假设迭代速度变快到当前常规水平的2倍，在3090显卡生成一个3D资产算力花费约2.6元左右。

文生3D资产算力成本远远低于市场3D资产购买价格，具有价格优势。在3D模型的在线内容库sketchfab上，即使简单难度的3D资产所需要的时间在2-15小时左右，购买费用在3-40美元区间，远远大于AI+3D生成成本。

图表:A100训练下需要的算力成本计算

情景	每秒迭代次数 (次/s)	单场景不同迭代次数下所需算力成本（元）			
		25000	30000	35000	40000
当前速度下	2.00	4.83	5.79	6.76	7.72
	2.25	4.29	5.15	6.01	6.86
	2.50	3.86	4.63	5.41	6.18
速度变快2X	4.00	2.41	2.90	3.38	3.86
	4.50	2.15	2.57	3.00	3.43
	5.00	1.93	2.32	2.70	3.09

图表:sketchfab平台售卖3D资产信息

3D模型复杂性	大概费用	所需时间 (小时)
最便宜	3美元	>2
简单	<40美元	>2
中等	<180美元	>15
复杂	<650美元	>40
非常复杂	>650美元	>50

五、投资建议：

3D对VR/AR游戏品类的影响：MR游戏，建议关注宝通科技、恺英网络、恒信东方；

3D对美术密集型品类的影响：开放世界游戏，建议关注网易；动画产业，关注光线传媒、奥飞娱乐；

3D对一般3D游戏公司的影响：3D游戏，建议关注2022年三季度低基数的三七互娱、姚记科技、凯撒文化、星辉娱乐；其他优质游戏公司如吉比特、神州泰岳、盛天网络、巨人网络、游族网络、完美世界。

六、风险提示：

技术发展不及预期，产业推进不及预期，法律及道德风险。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

本研究报告由方正证券制作及在中国（香港和澳门特别行政区、台湾省除外）发布。根据《证券期货投资者适当性管理办法》，本报告内容仅供我公司适当性评级为C3及以上等级的投资者使用，本公司不会因接收人收到本报告而视其为本公司的当然客户。若您并非前述等级的投资者，为保证服务质量、控制风险，请勿订阅本报告中的信息，本资料难以设置访问权限，若给您造成不便，敬请谅解。

在任何情况下，本报告的内容不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求，方正证券不对任何人因使用本报告所载任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告版权仅为方正证券所有，本公司对本报告保留一切法律权利。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。如需引用、刊发或转载本报告，需注明出处且不得进行任何有悖原意的引用、删节和修改。

类别	评级	说明
公司评级	强烈推荐	分析师预测未来12个月内相对同期基准指数有20%以上的涨幅。
	推荐	分析师预测未来12个月内相对同期基准指数有10%以上的涨幅。
	中性	分析师预测未来12个月内相对同期基准指数在-10%和10%之间波动。
	减持	分析师预测未来12个月内相对同期基准指数有10%以上的跌幅。
行业评级	推荐	分析师预测未来12个月内行业表现强于同期基准指数。
	中性	分析师预测未来12个月内行业表现与同期基准指数持平。
	减持	分析师预测未来12个月内行业表现弱于同期基准指数。
基准指数说明		A 股市场以沪深300 指数为基准；香港市场以恒生指数为基准，美股市场以标普500指数为基。



方正证券研究所

专注 · 专心 · 专业

北京市西城区展览路48号新联写字楼6层
长沙市天心区湘江中路二段36号华远国际中心37层

上海市静安区延平路71号延平大厦2楼
深圳市福田区竹子林紫竹七道光大银行大厦31层

广州市天河区兴盛路12号楼隼峰院2期3层方正证券