

证券研究报告
互联网传媒行业 / 行业深度报告
2023年4月13日

游戏的AI革命：StableDiffusion到Nerf

分析师：杨晓峰

登记编号：S1220522040001

方正证券（601901.SH）是行业领先的大型综合类证券公司，致力于为客户提供交易、投融资、财富管理等全方位金融服务。
Founder Securities (601901.SH), an industry-leading large comprehensive securities company, is committed to providing its clients with full services in stock transactions, investment & financing, wealth management, among others.

1、AI生成2D作画：Stable Diffusion。Stable Diffusion是一个文本到图像的潜在扩散模型，可以实现“文生图”功能。当前的输入内容类似于程序代码，具有一定的门槛。类GPT大语言模型可生成 prompt，进一步提升效率。目前 Stable Diffusion已经开源，并可生成多种风格图片；同时，使用 controlnet插件可以生成多视角图片。

2、AI推动“2D到3D”：NeRF模型。NeRF，全称为Neural Radiance Fields(神经辐射场)，是一项利用多视角图像重建三维场景的技术。英伟达的 instant NGP可以极大的提高NeRF的效率；龙头公司Luma AI，开发了NeRF相关的APP，目前已上线APP Store，极大的降低了NeRF的使用门槛，Luma AI近期已获英伟达投资。

3、从GPT→Stable Diffusion→Nerf：无中生有做3D。目前由文字到3D模型，存在两种路径。1)使用“文生图”模型+NeRF生成3D模型，GPT在其中可以发挥重要作用。例如，使用stable diffusion+controlnet插件，生成多视角的 2D图，再由NeRF形成3D建模。但由于目前文生图模型的 prompt仍有一定的门槛，因此或可训练 GPT，使其掌握，进而实现自然语言生成满意的多视角 2D图。2)使用文字生成3D模型。这一类模型往往是，文本到图像扩散模型和NeRF的结合。其原理为，先通过文本到图像的扩散模型生成2D图，再通过NeRF将2D图生成3D模型。

4、游戏制作流程：AI如何推动降本增效。当前制作一个3D游戏角色的周期在30-45天左右，且需要的步骤较多，参与的人员众多；使用 NeRF可以快速建模，从而提高效率，节省成本。

5、游戏美术岗位拆解：职责&成本。游戏美术的岗位众多，参与的人数也多，薪资平均在20K-35K/月，部分会达到30-50K/月。同时，外包美术的价格同样不低，大团队通常在 1500元/人/天。

6、投资建议：

1)我们认为 NeRF技术将会为游戏行业带来降本增效。建议关注：三七互娱、吉比特、盛天网络 姚记科技、恺英网络 神州泰岳、宝通科技、汤姆猫、天娱数科、巨人网络和完美世界。

2)动画行业或能受益于美术技术的进步，从而提升产能。建议关注：奥飞娱乐、光线传媒、上海电影。

7、风险提示：技术进步不及预期、玩家技术效果不认可、道德风险。

- 一、AI生成2D作画：Stable Diffusion
- 二、AI推动“2D到3D”：NeRF模型
- 三、从GPT→Stable Diffusion→Nerf：无中生有做3D
- 四、游戏制作流程：AI如何推动降本增效
- 五、游戏美术岗位拆解：职责&成本
- 六、投资建议及风险提示

行业路径的变革进度

游戏 美术制作流程	美术风格 设计	人物&场景 2D示意图	多视角 2D示意图	3D模型
AI可替代性	创意工作 仍需人工接入	text-to-image AI模型	视角定位插件	2D-to-3D AI模型
AI模型/算法		Stable Diffusion Dalle-2等	ControlNet	NeRF算法
原流程成本构成	10%	30%		60%

一、AI生成2D作画：Stable Diffusion

1 文生图模型：Stable Diffusion

Stable Diffusion是一个文本到图像的潜在扩散模型（**Diffusion**）。使用这个模型，可以生成包括人脸在内的任何图像。

由于Stable Diffusion提供了开源的预训练模型，用户可自行为其进行额外的训练，当前在插件生态上优于其他文生图模型。

图表：部分text-to-image模型

公司	模型	是否开源/开放使用
Stability AI Runway	Stable Diffusion	开放使用且开源
Open AI	Dalle Dalle2	开放使用
Google	Imagen Phenaki	否
Meta	Make-A-Video	否

1.1 文本：Stable Diffusion文本输入或可实现自动化

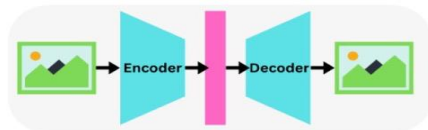
Stable Diffusion当前的输入内容类似于程序代码。需要分别输入正向 prompt（希望图像里出现什么）和负向prompt（不希望图像中出现什么），并且通过调整 prompt顺序，“{}”中括号强调，“: + 数字”赋值的方式调整权重。

类ChatGPT等大语言模型可生成Prompt，进一步提高生成效率。ChatGPT可在学习Stable Diffusion语法后根据用户的描述性语言生成相应的Prompt，从而进一步降低文生图模型的使用门槛和生成效率。

图表：Stable Diffusion模型原理

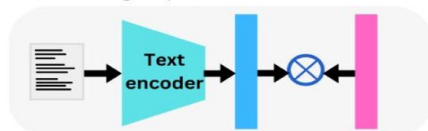
Step 1 Train a Variational Autoencoder

Create a latent space representation of the image



Step 2 Train a CLIP model to encode Text

Create a text embedding semantically aligned with the image representation



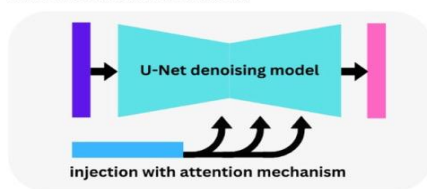
Step 3 Apply forward diffusion

Iteratively apply random noise to the latent representation of the image



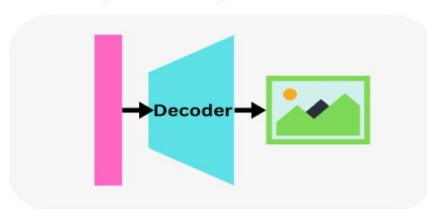
Step 4 Train a Denoising model with text prompt

Learn to denoise the diffused image embedding and inject the text embedding with attention mechanism



Step 5 Decode the image

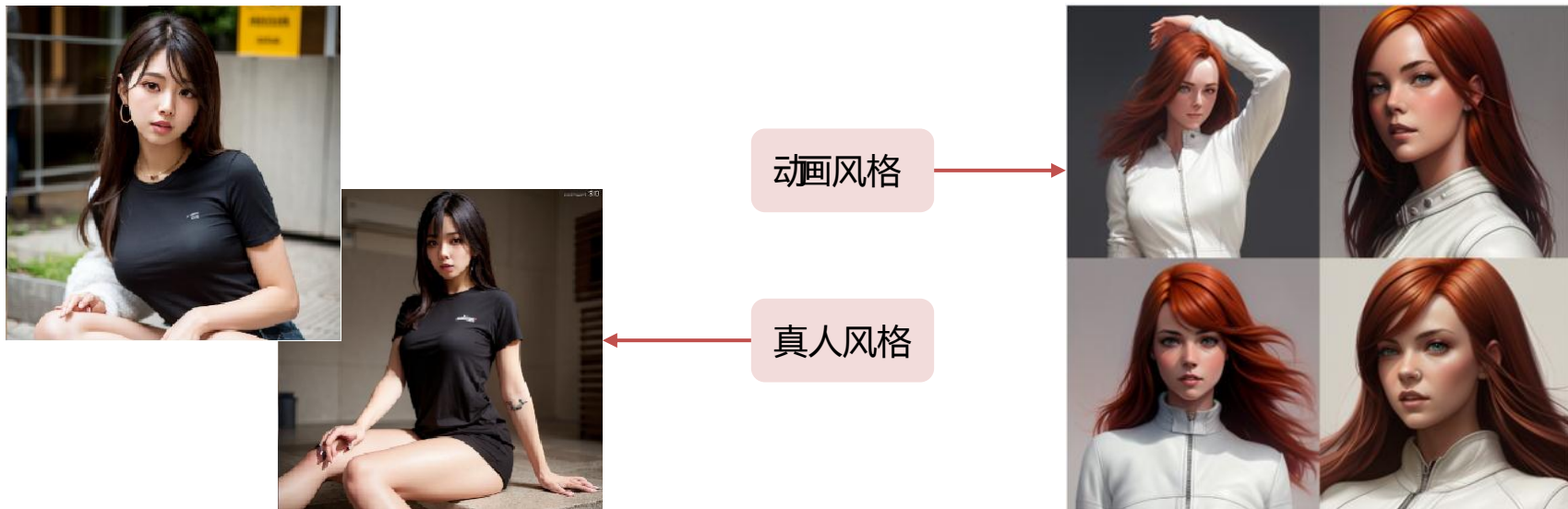
Use the trained autoencoder to decode the embedding into an image



1.2 图像：输出风格多样化

Stable Diffusion提供开源模型，用户可对其进行进一步训练，风格多样化。Stable Diffusion模型出现之初以生成二次元画风和拟真人物为主。后续通过用户训练为提供更多可选的预训练模型模版，生成的图像风格更加多样。

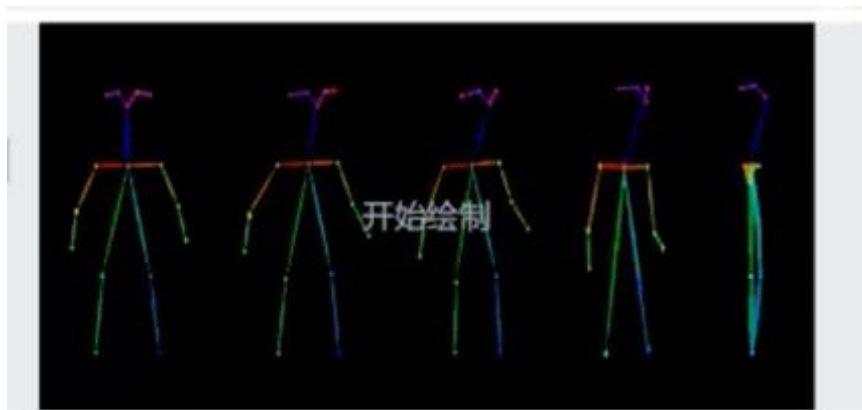
图表：Stable Diffusion 生成图片



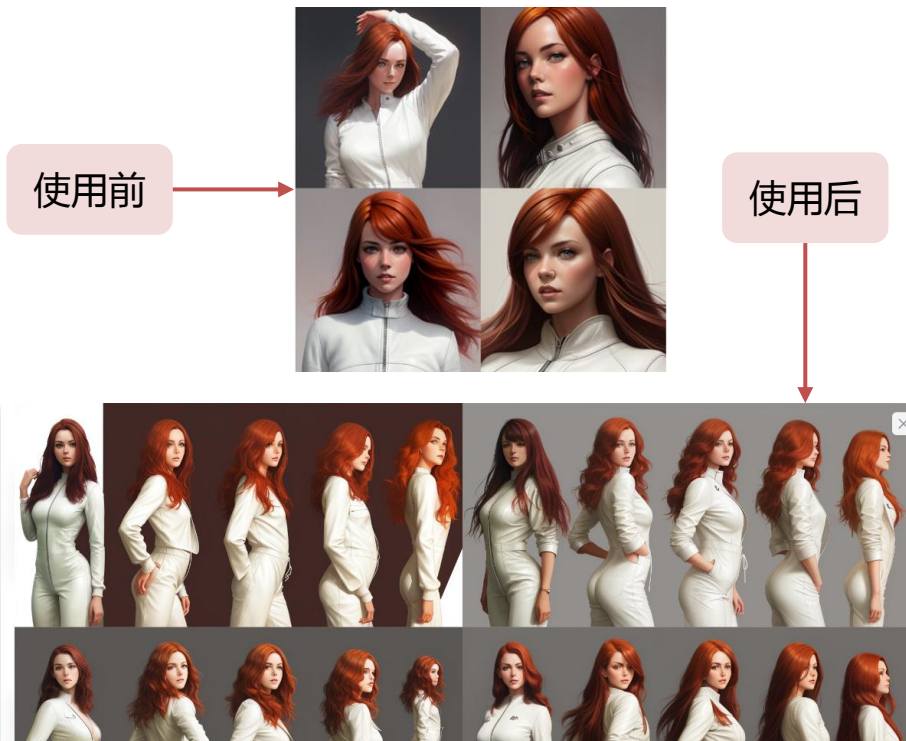
2、ControlNet插件：使Stable Diffusion可生成多视角图片

ControlNet插件为Stable Diffusion生成多视角图片提供可行方案。ControlNet通过在Stable Diffusion中确定人物的骨骼位置，从而生成人物相应的多视角图片，实现输出图像的多视角化。

图表：使用ControlNet插件插入视角图



图表：使用视角图前后对比

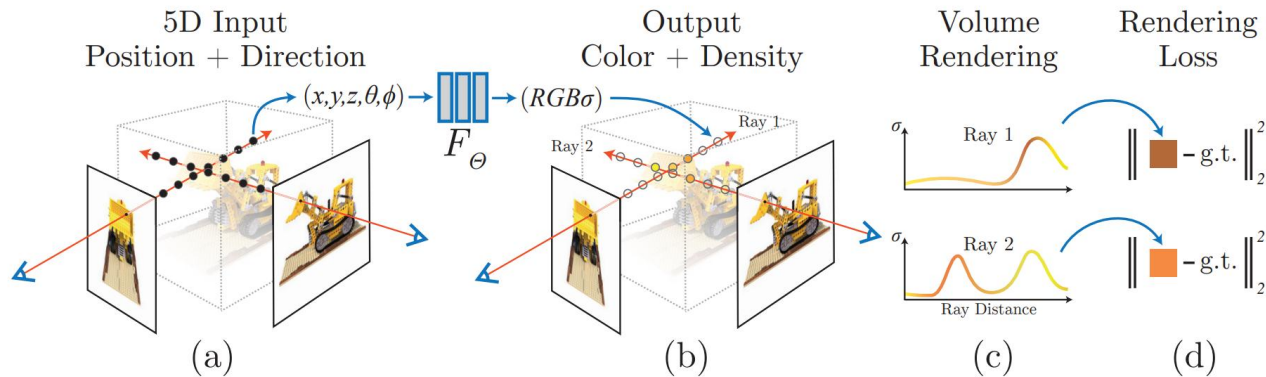


二、 AI推动 “ 2D到3D” : NeRF模型

1 NeRF : 从2D到3D的神经辐射场模型

- **NeRF**，全称为**Neural Radiance Fields(神经辐射场)**，是一项利用多视角图像重建三维场景的技术，由加州大学伯克利分校，Google研究院，及加州大学圣地亚哥分校的Ben Mildenhall等人在2020年提出。
- **实现NeRF的过程：**
 - 1、通过分析照相视角射线，从一组图片中生成一组采样点；
 - 2、将获得的采样点以及与之对应的 2D视角方向作为输入，输出一组颜色和体素；
 - 3、利用体素渲染技术和之前得到的颜色和密度生成希望看到的任意视角照片。

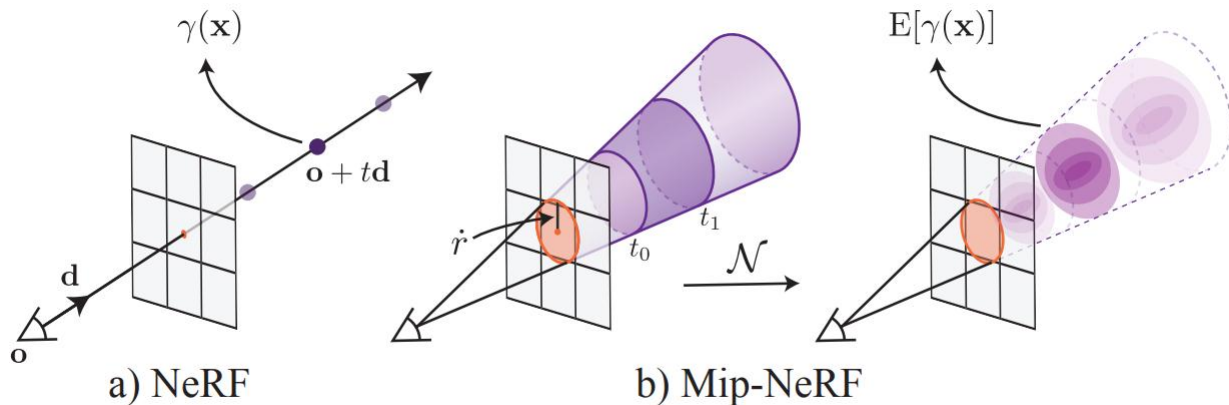
图表：NeRF模型的优化过程



1.1 Mip-NeRF：在NeRF的基础上提升建模效果

Mip-NeRF优化场景采样方式。 NeRF在渲染时使用每像素的单射线对场景进行采样，当训练或测试图像分辨率不同时，可能产生模糊或混淆的渲染效果。 Mip-NeRF扩展了NeRF，以连续值尺度表示场景。通过高效地渲染反锯齿的圆锥体而不是射线， Mip-NeRF减少了混叠伪影，并显示提高了 NeRF的细节表达能力，同时也比NeRF快7%，规模减半。与 NeRF相比，mip-NeRF在以NeRF呈现的数据集上降低了17%的平均错误率。

图表：Mip-NeRF模型较NeRF采样方式的改进



1.2 Block NeRF: 扩大NeRF的生成场景规模

Block NeRF扩大生成场景规模。 Google发布的《Block-NeRF：可扩展的大场景神经视图合成》的最新研究成果，重点就是将NeRF的应用场景从小微场景或者说单个场景对象扩展到的城市级别，并实现了在场景需要更新的时候只需要进行场景的分块更新，而不再需要整个场景的全量训练和更新。

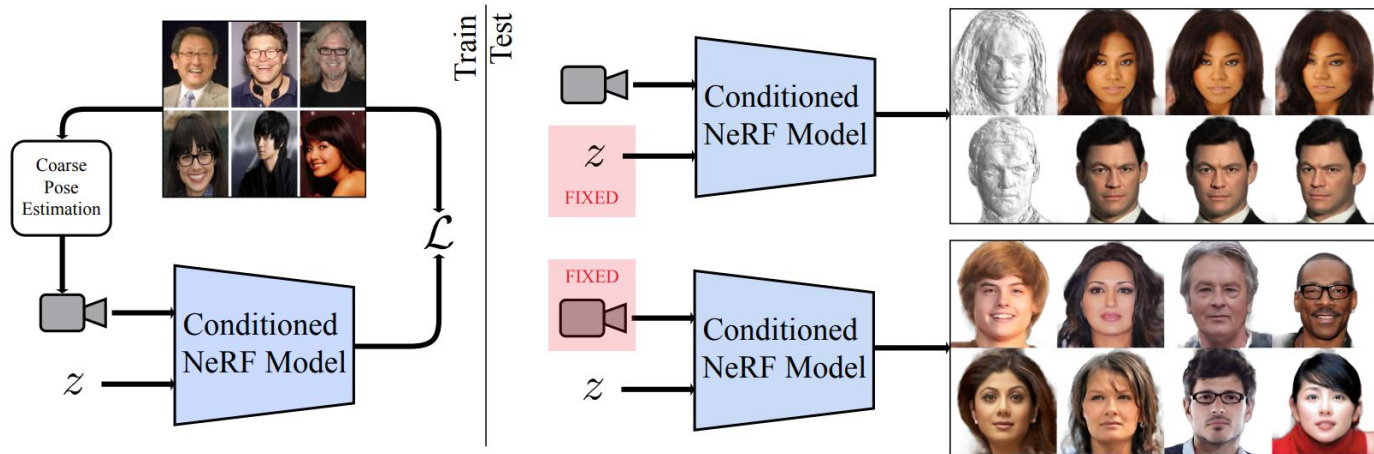
图表：Block NeRF模型生成的街区3D模型



1.3 LOLNeRF : 降低NeRF对图片数量的要求

LOLNeRF降低输入图片数量要求。 来自英属哥伦比亚大学，西蒙菲莎大学和Google Research的研究人员发提出了LOLNeRF，对于同一类物体来说，仅需单一视角即可训练 NeRF模型，而无需对抗监督。一旦共享的生成模型训练完毕，模型即可提供近似的相机姿态（camera poses）。LOLNeRF使用预测的二维 landmarks将数据集集中的所有图像大致对齐到一个典型的姿态，以此来确定应该从哪个视图渲染辐射场以再现原始图像。

图表：LOLNeRF的运行逻辑

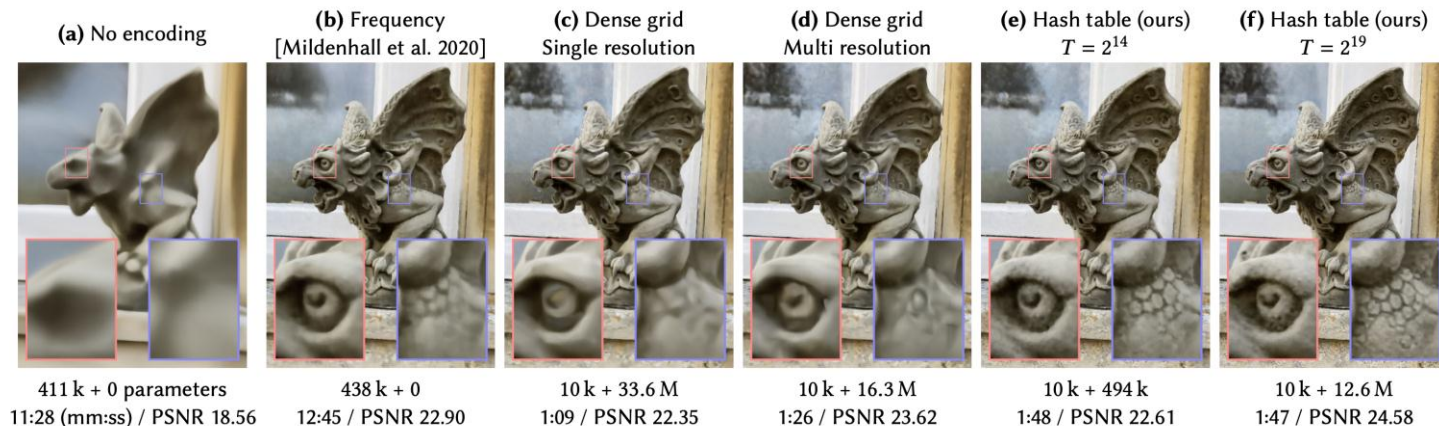


2 Instant NGP : 极大提高NeRF的生成效率

英伟达NVIDIA Research团队开发出一种方法，几乎能在瞬间内完成这项任务，是同类中首批将超高速神经网络训练与快速渲染相结合的模型之一。英伟达将该方法应用在 NeRF技术上，创造出Instant NeRF在某些情况下速度可以提升超过1000倍。

Instant NeRF可在单GPU (RTX3090) 环境下实现秒级的3D场景生成。在高清分辨率下，合成甚至真实场景可以在几秒内训练，并以 60帧/秒的速度渲染。

图表：Instant NeRF的训练效果 (图e、f)



3.1 Nerf独角兽—luma：极大的降低了NeRF的使用门槛

- **Luma为一款手机APP，已上架APP Store，支持iPhone 11或以上的机型，iOS16.0及以上版本，应用大小为40M。**由Luma AI团队基于云服务器开发。它极大的降低了NeRF的使用门槛，仅需要一部手机，通过APP导引进行场景的拍摄，就可以随时随地渲染三维场景，轻松使用NeRF。而以往的大多数NeRF衍生算法，需要本地部署运行环境，对配置要求较高，算力消耗大。Luma AI把训练和渲染步骤搬到云端服务器，从而降低了使用门槛。
- **Luma渲染效果大幅提升且适用场景更加广泛。**其它NeRF衍生算法的适用的场景有限，有些适用于大场景，有些在小场景中表现更佳，无法满足普适化的需求。Luma AI在算法上进行了优化和改进，适用各种场景，小到各种小摆件大到城市街区等场景，渲染效果更加逼真。
- **Luma支持导出渲染场景Mesh模型和点云模型，**有利于进一步扩展应用。

图表：Luma通过视频（左）生成的3D模型（右）



3.2 Nerf独角兽—luma：近期获英伟达投资

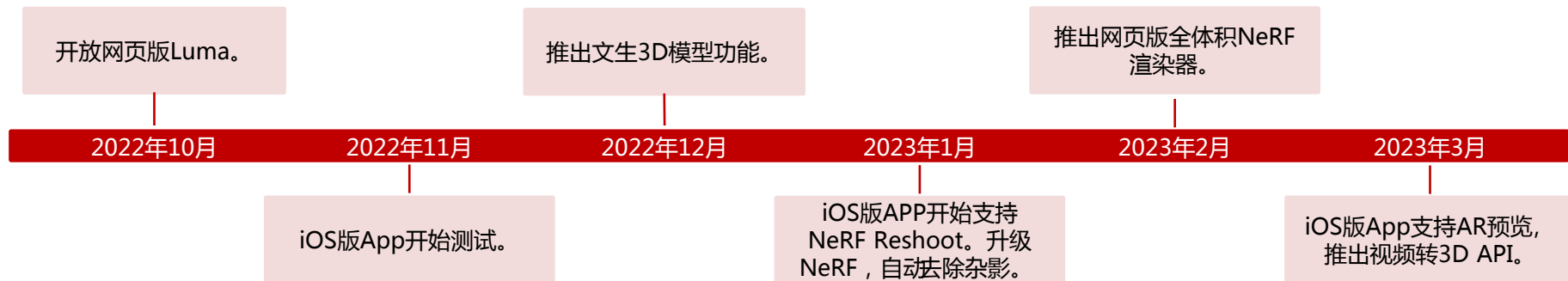
2023年3月24日，Luma获2000万美元的A轮融资，英伟达为其投资方之一。

2023年3月，Luma推出视频转3D场景API：Video-to-3D API，进一步开放 NeRF的能力，并进行商业化探索。

图表：Luma融资情况

披露日期	融资轮次	交易金额	投资方
2023-03-24	A轮	2000万美元	Amplify Partners；General Catalyst；Matrix Partners；South Park Commons；英伟达
2021-10-30	种子轮	430万美元	South Park Commons；Matrix Partners；Context Ventures；Amplify Partners；个人投资者

图表：Luma近半年发展历程



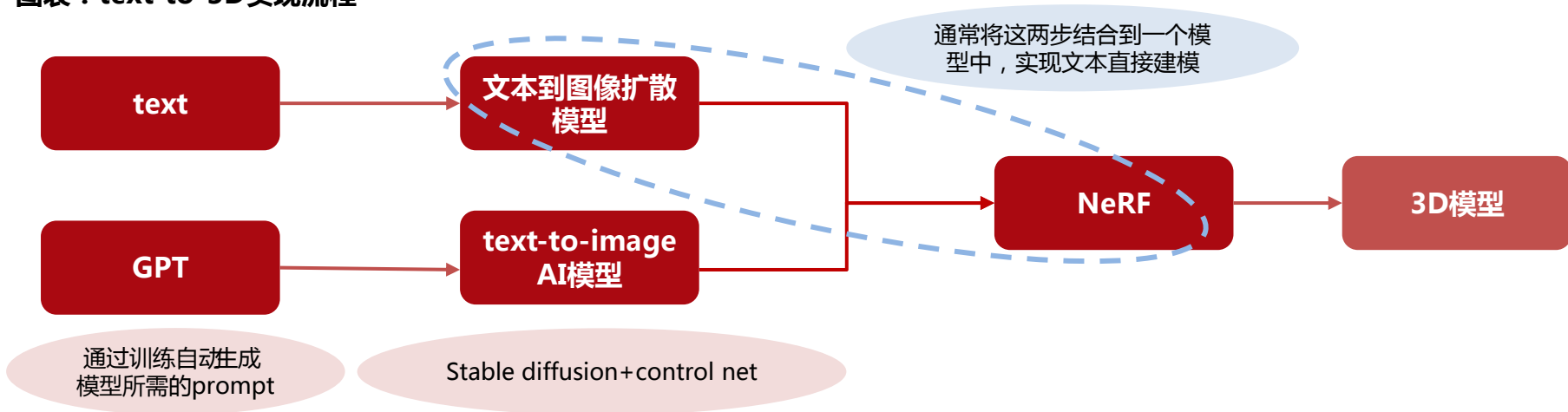
三、从GPT→Stable Diffusion→Nerf : 无中生有做3D

1 无中生有做3D：一般有两种路径

➤ 目前由文字到3D模型，存在两种路径

- 使用“文生图”模型+NeRF生成3D模型，GPT在其中可以发挥重要作用。例如，使用stable diffusion+controlnet插件，生成多视角的 2D图，再由NeRF形成3D建模。但由于目前文生图模型的prompt仍有一定的门槛，因此或可训练 GPT，使其掌握，进而实现自然语言生成满意的多视角 2D图。
- 使用文字生成3D模型。这一类模型往往是，文本到图像扩散模型和NeRF的结合。其原理为，先通过文本到图像的扩散模型生成2D图，再通过NeRF将2D图生成3D模型。

图表：text-to-3D实现流程

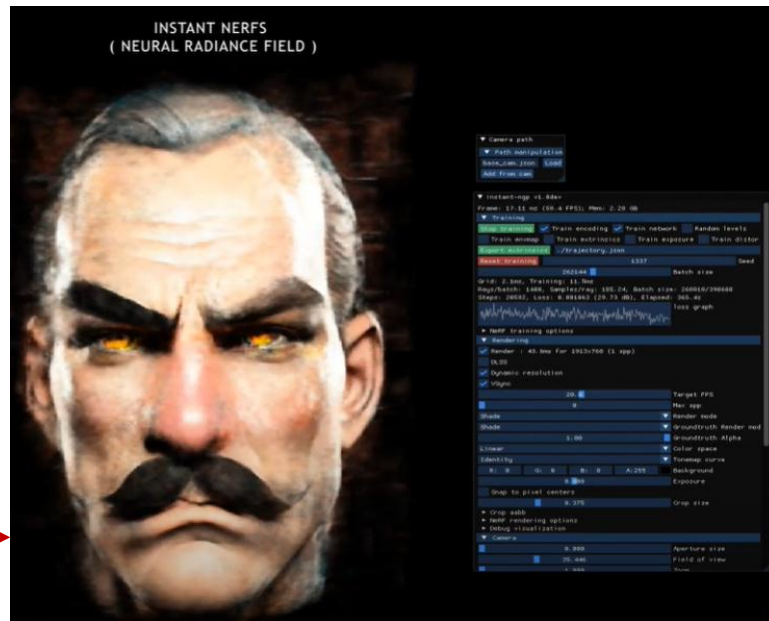


2 实例：Stable Diffusion→NeRF



用Stable Diffusion获得了
3张不同视角的头部图片

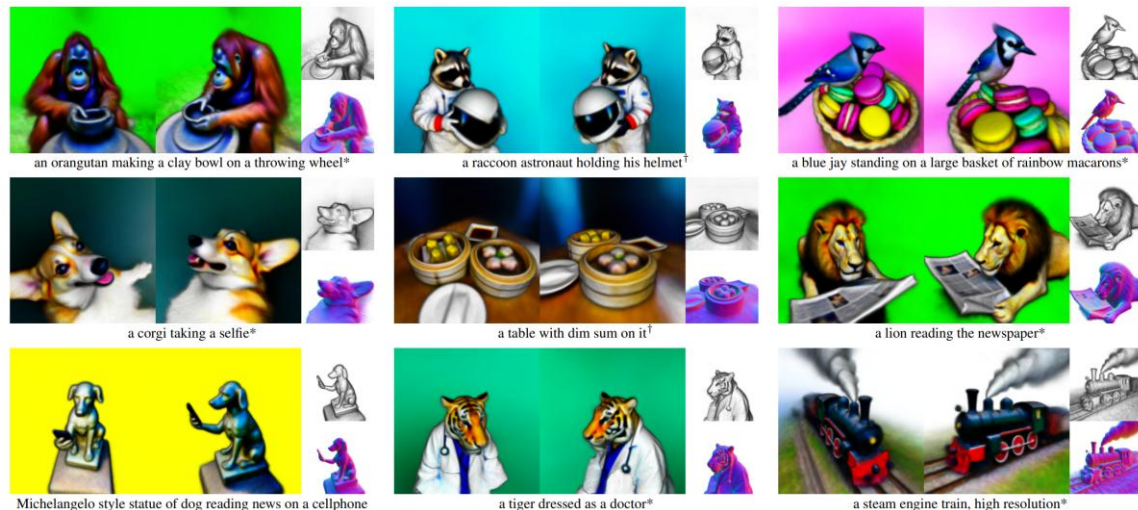
通过NeRF获得3D模型



3.1 实例—文字生成3D：Dream Fusion

Google研究员提出新模型DreamFusion，先使用一个预训练2D扩散模型基于文本提示生成一张二维图像，然后引入一个基于概率密度蒸馏的损失函数，通过梯度下降法优化一个随机初始化的神经辐射场NeRF模型。训练后的模型可以在任意角度、任意光照条件、任意三维环境中基于给定的文本提示生成模型，整个过程既不需要3D训练数据，也无需修改图像扩散模型，完全依赖预训练扩散模型。

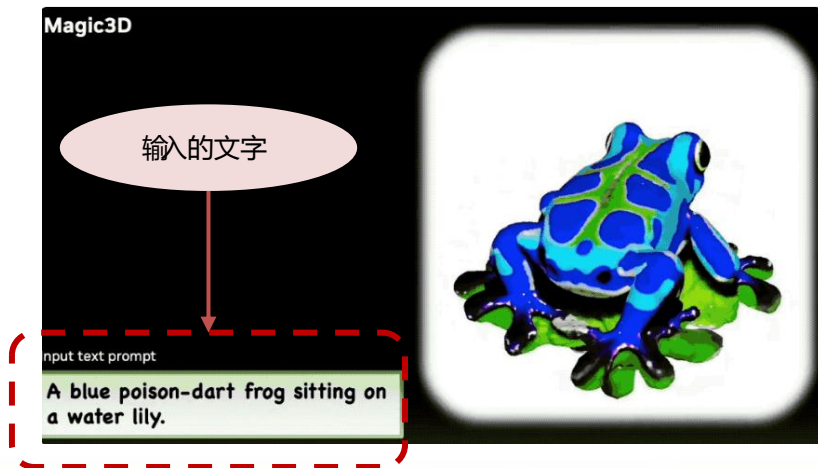
图表：DreamFusion生成的部分3D模型



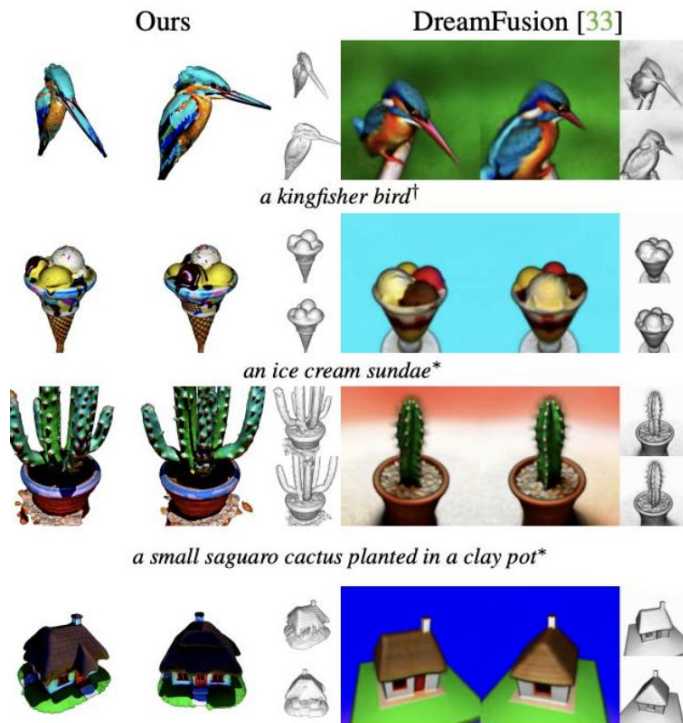
3.2 实例—文字生成3D：英伟达Magic3D AI

- Magic 3D AI生成3D模型，先将文字扩散至图像，随后使用instant NeRF形成3D模型。
- Magic 3D AI生成的模型比DreamFusion分辨率高8倍，速度快2倍，只需40分钟即可完成渲染。但目前尚未开源。

图表：Magic3D AI生成的部分3D模型



图表：Magic3D AI与Dreamfusion对比



四、游戏制作流程：AI如何推动降本增效

1 游戏美术的四种制作方式



目前，国内常见的游戏美术制作方式有四种，分别是3渲2制作、3D 现世代制作（传统手绘）、3D 传统次世代制作和3D次世代制作。

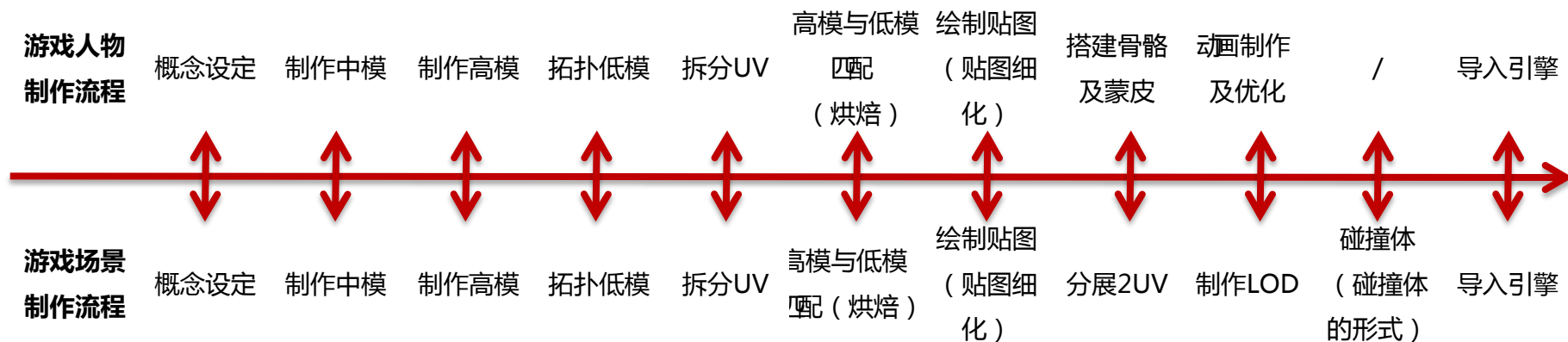
图表：游戏制作方式分类

制作方式	代表作	特征点
3渲2制作	《梦幻西游》系列、《大话西游》系列、《率土之滨》系列等	①2D画面精致,②游戏客户端包体小,③非硬核玩家要求友好,④有固定用户群
3D 现世代制作 (传统手绘)	《天下3》	①重要的艺术风格,②比画面技术实力重要许多,③风格突出，画面统一
3D 传统次世代制作	《剑灵》、《天刀OL》	①高画质、高品质、高细节 ,②使用高端游戏引擎制作,③高模、贴图多
3D 次世代制作	《神都夜行录》、《楚留香》	①真实度越来越高,②精准定义材质属性,③PBR (Physically Based Rendering)

2 当前制作流程：步骤较多、耗时较长

- 通常制作一个3D游戏角色的周期在30天到45天左右。
- 传统美术制作流程分为人物制作流程和场景制作流程，都包含概念设计，3D建模（中模-高模-低模），分展UV，烘焙，绘制贴图和导入引擎。
- 不同的是人物制作流程还需要搭建骨骼和蒙皮，动画制作及优化。场景制作流程还包括分展 2UV，制作 LOD和碰撞体。

图表：游戏人物/场景制作流程



3 NeRF推动美术降本增效：节省时间和人力

- 时间上，AI生成图片的速度相对手画来说，大幅度提高；而NeRF建模也可一次性完成上文中步骤的多个，在时间上有极大的提高。
- 人力上，“文生图”和NeRF可以完成多个角色的工作，同时也更加快速。

图表：当前美术制作流程与NeRF对比

当前流程		NeRF
人物	如上述步骤进行建模	使用“文生图”快速生成图片，之后使用 NeRF建模
虚拟场景	如上述步骤进行建模	
实景	如上述步骤进行建模	通过拍摄实景视频，使用NeRF建模

五、游戏美术岗位拆解：职责&成本

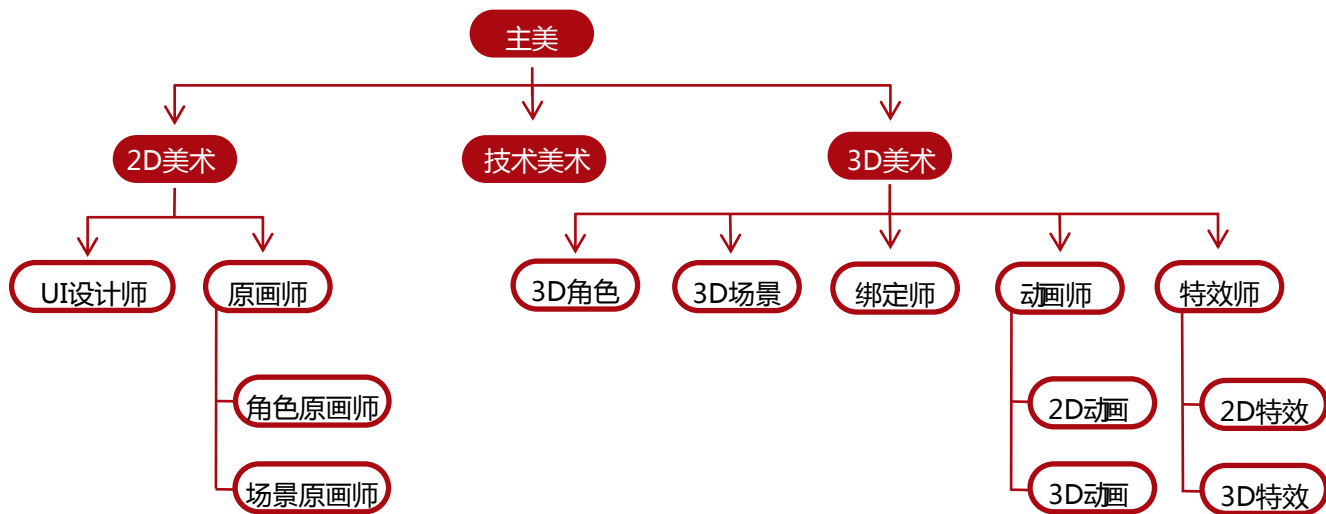
1 职责拆解：美术团队中角色众多，因此需要的人数也多

游戏美术团队由项目组主美术来掌握全局，把握美术的方向；往下具体分为2D美术、技术美术和3D美术。

2D美术包括UI设计师和原画师，原画师具体分为角色原画师和场景原画师。

3D美术包括3D角色、3D场景、绑定师、动画师和特效师，其中，动画师分 2D动画和 3D动画，特效师分 2D特效和3D特效。

图表：游戏美术团队的岗位构成



2 薪资：普遍在20-30K/月，部分在30-50K/月

从薪资上来看，游戏美术岗位的薪资区间跨度是较大的，但大多数人的薪资是比较可观的。

项目组主美术、3D角色设计师和3D场景设计师中的大部分人的薪资可以达到30K-50K元/月，UI设计师、原画师、特效师等则可以达到20K-30K元/月；这与游戏行业人才需求增加和薪资提升是密切相关的。

从需求上来看，在游戏研发团队中，2D岗位的需求量占总设计师人数的15-20%左右，3D岗位的需求量占总设计师人数的70-80%左右。总体来看，3D岗位的需求量是大于2D岗位的，3D岗位入手也比2D岗位入手要容易一些。

图表：游戏美术岗位的薪资区间和大多数人的薪资

岗位	薪资区间（元/月）	大部分人的薪资（元/月）
主美	2K - 50K	30K-50K
UI设计师	3K - 50K	20K-30K
角色原画师	4.5K - 50K	20K-30K
场景原画师	4.5K - 50K	20K-30K
3D角色设计师	4.5K - 50K	30K-50K
3D场景设计师	4.5K - 50K	30K-50K
2D动画 设计师	4.5K - 50K	20K-30K
3D动画 设计师	4.5K - 50K	10K-15K
2D特效设计师	3K - 50K	20K-30K
3D特效设计师	3K - 50K	20K-30K

3 业务外包：大团队外包报价在1500元/人/天

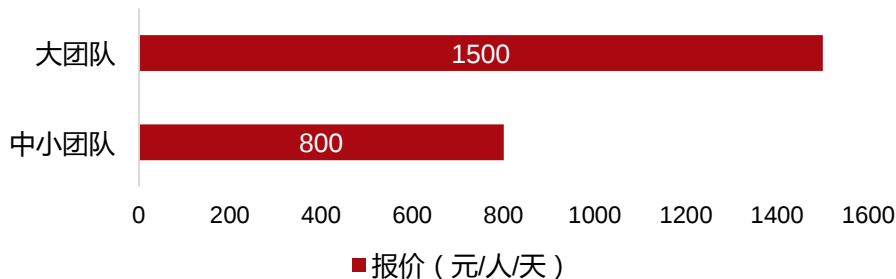
通常选择外包的业务：

- 1) 游戏的宣传动画和音乐等通常直接外包给指定的大触
- 2) 堆量的东西基本会选择外包来满足需求
- 3) 技术要求高的业务有时候也会选择外包

通常不会外包的业务：

- 1) 设计；2) 内容复杂且系统的业务

图表：大团队和中小团队业务外包的报价



图表：普通/次世代3D角色建模外包所需的人数和价格

类型	人数	价格
普通3D角色建模	5-8人/天	数千元
次世代3D角色建模	15-20人/天	上万元

六、投资建议及风险提示

投资建议：

- 1) 我们认为 NeRF技术将会为游戏行业带来降本增效。建议关注：三七互娱、吉比特、盛天网络 姚记科技、恺英网络 神州泰岳、宝通科技、汤姆猫、天娱数科、巨人网络和完美世界。
- 2) 动画行业或能受益于美术技术的进步，从而提升产能。建议关注：奥飞娱乐、光线传媒、上海电影。

风险提示：

技术进步不及预期、玩家技术效果不认可、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。

公司投资评级的说明

强烈推荐：分析师预测未来半年公司股价有20%以上的涨幅；
推荐：分析师预测未来半年公司股价有10%以上的涨幅；
中性：分析师预测未来半年公司股价在-10%和10%之间波动；
减持：分析师预测未来半年公司股价有10%以上的跌幅。

行业投资评级的说明

推荐：分析师预测未来半年行业表现强于沪深 300指数；
中性：分析师预测未来半年行业表现与沪深300指数持平；
减持：分析师预测未来半年行业表现弱于沪深 300指数。

专注 专心 专业



方正证券研究所

北京市 西城区展览路 48号新联写字楼6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 隼峰苑2期3层方正证券

长沙市 天心区湘江中路二段 36号华远国际中心37层

更多免费行业报告

并购家

www.ipoipo.cn