

人工智能生成内容在虚拟现实领域的应用研究

刘成坤¹, 李赛伽¹, 李振¹, 伍孟晗¹

(1. 北京理工大学 计算机学院, 北京 102400.)

摘要: 虚拟现实技术 (VR) 在计算机图形技术、多媒体技术、传感技术、实时技术、人工智能和仿真技术等方面的应用。虚拟现实技术通过将参与者置身于具有视觉、听觉、触觉和味觉等多种感官体验的三维虚拟环境中, 实现了身临其境的沉浸感。近年来, 基于大型语言模型的人工智能生成内容技术 (AIGC) 快速发展, 本文探讨了大语言模型的发展现状, 并论述了 AIGC 的巨大生产优势。同时结合前沿技术, 论述了 AIGC 在 VR 领域的模型生成、场景生成、虚拟对话生成的关键技术与现存的问题, 并展望 AIGC 在 VR 领域未来的发展前景。

关键词: 虚拟现实; 人工智能生成内容

中图分类号: BIT001

文献标志码: SSS

文章标号: 123

DOI: [1120200167/08012003](https://doi.org/10.11918/0013788X.2023.05.001)

Research on the Application of Artificial Intelligence Generated Content in the Virtual Reality Field

LIU Chengkun¹, LI Saijia¹, LI Zhen¹, WU Menghan¹

(1. School of Computer Science & Technology, Beijing Institute of Technology, Beijing 102400, China)

Abstract: Virtual reality technology (VR) finds its applications in computer graphics, multimedia technology, sensing technology, real-time technology, artificial intelligence, and simulation technology. VR technology immerses participants in a three-dimensional virtual environment that provides multisensory experiences such as visual, auditory, tactile, and gustatory sensations. In recent years, artificial intelligence-generated content (AIGC) based on large language models has rapidly developed. This paper explores the current development status of large language models and discusses the significant production advantages of AIGC. Additionally, this paper combines cutting-edge technologies to discuss the key techniques and existing issues of AIGC in the VR field, including model generation, scene generation, and virtual dialogue generation. Furthermore, it outlines the future development prospects of AIGC in the VR domain.

Key words: Virtual Reality; Artificial Intelligence Generated Content

收稿日期: 2023-06-11

基金项目: 自费

作者简介: 刘成坤(2001—), 男, 软件工程本科生, E-mail: ck167@bit.edu.cn ID: 1120200167

李赛伽(2001—), 男, 软件工程本科生, E-mail: 1120201450@bit.edu.cn ID: 1120201450

伍孟晗(2002—), 男, 软件工程本科生, E-mail: 927115185@qq.com ID: 1120200730

李振(2002—), 男, 软件工程本科生, E-mail: abtrvmuzi16@126.com ID: 1120200268

1. 绪论

1.1 虚拟现实技术概述

虚拟现实技术 (Virtual Reality, VR) 是 20 世纪 90 年代发展起来的一种以计算机技术为核心的新技术, 通过对计算机图形技术、多媒体技术、传感技术、实时技术、人工智能、仿真技术等结合, 使参与者置身于集视觉、听觉、触觉、味觉等于一体的三维虚拟环境中, 借助特殊的输入输出设备, 用自然的方式与虚拟世界中的物体进行交互, 产生沉浸感^[1]。为了实现这一目标, 虚拟现实系统就需要完善的 3D 模型和场景渲染, 需要使用计算机图形学技术来创建逼真的 3D 模型和场景, 并将其渲染到用户的显示设备上。这些模型和场景包括虚拟环境中的人物、物体、景观等, 以及与之相关的光照、阴影和纹理等效果。此外, 为了产生更加逼真的沉浸感, 还需要为交互对象增加更加真实的对话、故事背景等。

虚拟现实技术在各领域都有广泛的应用, 包括娱乐、游戏、教育、医疗、建筑设计、旅游等。它为用户提供了一种全新的交互方式, 能够创造出身临其境的感觉, 扩展了人们的感知和体验范围。虚拟现实技术结合人工智能生成内容 (AI Generated Content, AIGC) 具有极大的潜力进一步提升用户体验, 具有巨大的前景。AIGC 技术利用人工智能算法和机器学习模型, 能够生成高度逼真的虚拟对象、场景和人物。

随着技术的不断发展和改进, 虚拟现实技术将会越来越成熟和普及, 为人们带来更多的创新和乐趣。

1.2 人工智能生成内容核心内涵

AIGC (Artificial Intelligence Generate Content), 即人工智能生成内容。目前, 学界对其还无具体的概念界定。中国信息通信研究院和京东探索研究院发布的关于 AIGC 的白皮书中指出, 由于国内学界对 AIGC 理解为是继 PGC、UGC 之后的利用人工智能技术自动生成内容的新型生产方式, 在国际上对应的术语是 AI-generated Media 或 Synthetic Media, 因此, AIGC 的概念即“ AIGC 既是从内容生产者视角进行分类的一类内容, 又是一种内容生产方式, 还是用于内容自动化生成的一类技术集合”^[2]。

简单来说, AIGC 就是一种基于人工智能技术的内容生成方法。它利用自然语言处理、机器学习和深度学习等技术, 通过分析大量的文本数据, 能够生成自然流畅的文本内容, 包括文章、新闻、故事、对话等。

AIGC 是建立在强大的语言模型之上的, 该模型通过学习大规模的文本语料库, 掌握了语言的语法结构、词汇用法和上下文关系。它可以根据给定的输入, 自动生成相关的内容, 具有一定的创造性和逻辑性。

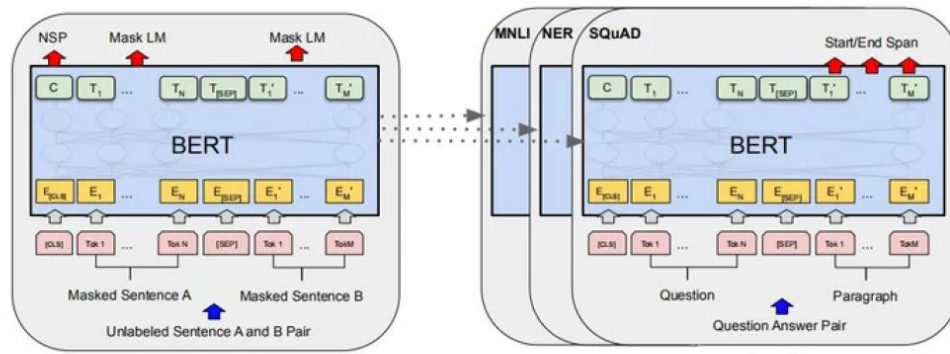


图 1 BERT 的训练过程: 预训练和模型微调

Fig.1 The training process of BERT: pre training and model fine-tuning

AIGC 的应用非常广泛。它可以用于自动写作, 帮助人们生成各种类型的文本, 如新闻报道、博客文章、市场营销材料等。它还可以用于辅助创作, 为作家提供灵感和构思, 填补故事情节的空白。

当 AIGC 与 VR 结合时, 人们能够体验到更加逼真和引人入胜的虚拟现实世界。通过 AIGC 的智能算法和深度学习技术, 虚拟场景中的角色、物体和环境能够呈现出更加逼真的行为和互动。无论是在游戏中探索奇幻的幻境, 还是在培训模拟中体验危险环境, AIGC 的应用使得虚拟现实变得更加真实而令人难以置信。

针对特定场景下和具体任务对模型进行微调。在语言模型中, 最为突出的两种模型包括 BERT 和 GPT^[3]。

2.1 BERT

BERT 是 Google 研究人员于 2018 年推出的语言表示模型。BERT 的训练过程分为两步: 预训练和模型微调, 如图 1 所示。

与先前存在的语言表征模型不同的是, 研究者们证明了双向与训练对于语言表征的重要性, 并且在预训练中令 BERT 采用了该种方式并且使用掩码方法实现预训练的深度双向表征, 在许多的标记任务上取得了最好的性能^[4]。

2. 大型语言模型

大语言模型 (Large language models, LLM) 是深度学习模型中的一个特殊模型, 目前已经成为计算机处理文本, 从文本分类到文本生成的最先进的解决方案。通常而言, 此类模型首先在大语言语料库上进行预训练, 也会

2.2 GPT

GPT 模型于 2018 年由美国 AI 组织 OpenAI 推出, 它是基于 Transformer 架构的人工神经网络。GPT 模型在未标记文本的大型数据集上进行预训练, 并且能够生成类似人类的内容。目前, 多数

LLM 都已经具备以上特征。

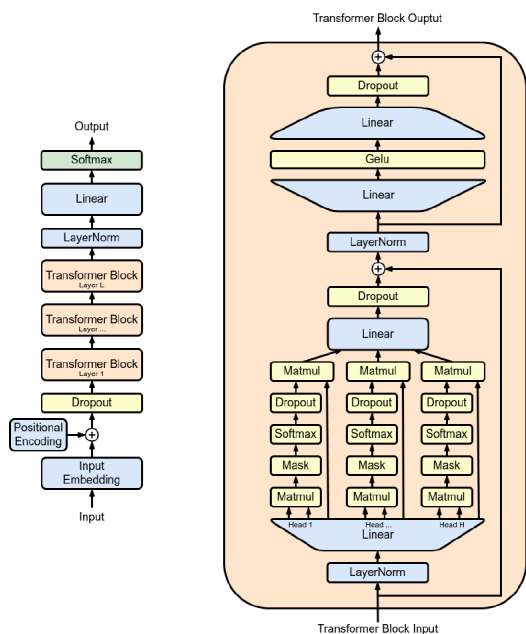


图 2 GPT 的训练过程

Fig.2 The training process of GPT:

作为对绝大多数的 LLM 都产生了极大影响的语言模型，GPT 的工作原理成为 AIGC 领域中的经典。GPT 基于一个拥有单个解码器的 Transformer 网络，模型在一个大型文本数据语料库中进行预训练，随后会针对特定的任务进行微调以使得它能够更好地胜任文本生成、问题的回答等工作。GPT 模型根据先前的标记来预测序列中的下一个标记，从而使得 GPT 模型获得了学习语言中的模式，以及单词与其意义的关系。

由 OpenAI 发布的 Chat GPT 广泛得到了大众的关注，使得大众得以认识到 Chat GPT 在理解自然语言和生成对话方面已经能够拥有人类的能力。近期公布的 GPT-4 更是表现出了相比以往的人工智能模型，具有更加普遍的智能。研究

人员证明，除了对于语言的掌握能力之外，GPT-4 还能够解决新的困难任务，范围涵盖数学、编码、视觉、医学、法律、心理学等等而不需要任何明显的提示，并且能够获得非常接近人类水平的表现 [5]。

3. AIGC 的关键优势

AIGC 的 AI 独特技术支撑使其具备生产力的强大、生产过程的个性化以及生产结果的优质可体验等独特优势 [6]。

3.1 强大的生产力

AI 技术从浅神经网络开始就具备了自学和运算能力，这使得 AI 技术支撑下的内容生产模式具有强大的内容生产力。主要体现在：

(1) 生产的全天候。AIGC 由算法产生劳动力，这就跟有限的人力形成对比。只要算法可以运行，AIGC 就可以实现 24 小时全天候进行内容生产。

(2) 生产的快速性。由于 AI 技术强大的运算能力，可将人脑认为复杂的创作，在短时间内通过机器学习快速转化成大量数据，完成创作，实现时间和数量上的飞跃。

(3) 生产的自主性。AIGC 的生产具有模拟人脑的功能且具有深度自学能力，让它不依赖于人的思维，而是独立的机器人。

(4) 生产的多样性。AIGC 并不局限于某一领域的创作，它既涵盖了文

学、管理学、传播学等人文社会学科,也被应用到生物学、科学等自然科学学科,具有明显的生产多样性。

3.2 个性化的生产过程

与传统的 UGC (User Generated Content) 生产模式类似, AIGC 在生产过程中也具有个性化的特征。但不同的是, UGC 是由多主体的用户参与实现个性化生产, 但 UGC 的内容生产需借助互联网平台的社交媒体为载体, 这就不可避免地受到平台中的群体压力、群体趋同等传播效应的影响, 容易出现内容同质化现象, 尤其是抖音、快手等平台经常出现一哄而上追逐当前热点的行为, 因此, 虽然是具有完全个性化的主体, 但并不能保证内容的完全个性化。而 AIGC 并非追求生产过程中用户的作用, 而是通过自动识别场景, 抓取数据, 寻找不同模态间的对应关系, 实现智能的个性化, 这就有效避免了同质化的现象。

3.3 优质可体验的生产结果

不同于 UGC、PUGC 中由于用户内容生产水平不同导致既有优秀的内容也存在低俗的成分, AIGC 由于自主和个性化的生产极大提高了生产内容的质量。其一, 由于算法的不断优化, 内容生产中低级的错误如语法、错别字等可以被自动检索并规避, 并且在内容语义

上可以模拟更高素养的人脑思维, 起到比人工审核更精准的内容把关; 其二, GAN 等技术支持了 AIGC 更高清的图像, 带来更好的视觉呈现。

除了内容优质, 另外 AIGC 通过构建多维数据, 将平面的场景转换成立体的模型, 促进了 2D 内容到 3D 内容的过渡, 还可以与 VR、AR 技术结合带来三维沉浸式的体验效果。

4. 三维模型生成

三维模型生成是一个长期研究且比较重要的问题, 计算机视觉、计算机图形学和机器学习社区已经对其进行了几十年的探索。自 2012 年以来, 卷积神经网络(CNN)对图像的处理能力优于很多传统方法。因此众多科研工作者也将卷积网络深度学习用来解决基于图像的三维重建问题, 并显示出很好的性能。从 2022 年的 AI 爆发以来, 利用 AI 自动建模也进入了大众视野, 随着 AI 领域新技术的不断发展, 基于 AI 的模型生成技术正在逐渐成为三维模型生成的新趋势^[7]。

4.1 AIGC 三维模型生成训练

生成模型训练是实现 AIGC 技术的关键。在三维模型生成领域, 生成模型训练通常采用深度学习方法。常见的深度学习模型包括生成对抗网络 GAN、变分自编码器 VAE 等^[8]。这些模型可

以通过大量的三维数据来学习三维模型的特征和规律,从而实现生成高质量逼真的三维模型。

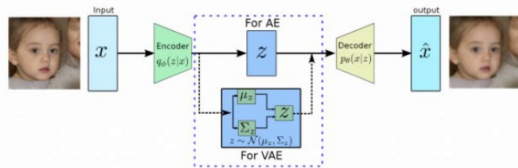


图 3 变分自编码器结构图

Fig.3 The struct of Variational Autoencoder

数据集构建是生成模型训练的基础。构建高质量的三维数据集是非常耗费成本,但是对于训练高质量的生成模型至关重要。数据集的构建需要考虑数据集的规模、多样性和真实性等方面,同时还需要注意数据集的质量和完整性,保证模型的训练效果。

模型优化对模型训练也具有非常重要的作用。通过算法调整和参数更新等手段,提高生成模型的性能和效率。在三维模型生成领域,模型优化可以采用多种方法,如基于梯度的优化方法、基于遗传算法的优化方法等。优化方法的

选择需要根据具体情况进行调整,以实现更高效、更准确的三维模型生成。

4.2 AIGC 模型生成前沿探索

AIGC 在过去几年获得了显著成就,文本引导的内容生成是最实用的一种^[9], Spline AI 就是其中之一。Spline AI 是目前最为知名的通过文字生成三维模型的人工智能模型,它可以帮助用户快速生成高质量的三维模型。该工具的核心技术是基于深度学习的生成模型,可以在不需要手动建模的情况下,自动生成符合用户需求的三维模型。

用户可以通过输入一些简单的几何图形,如球体、立方体、圆柱体等,来描述他们需要生成的三维模型的基本形状, Spline AI 会使用深度学习模型来生成符合用户需求的三维模型,用户可以使用 Spline AI 的编辑工具对生成的模型进行微调和优化以达到更好的效果。

Spline AI 通过学习来自真实三维模

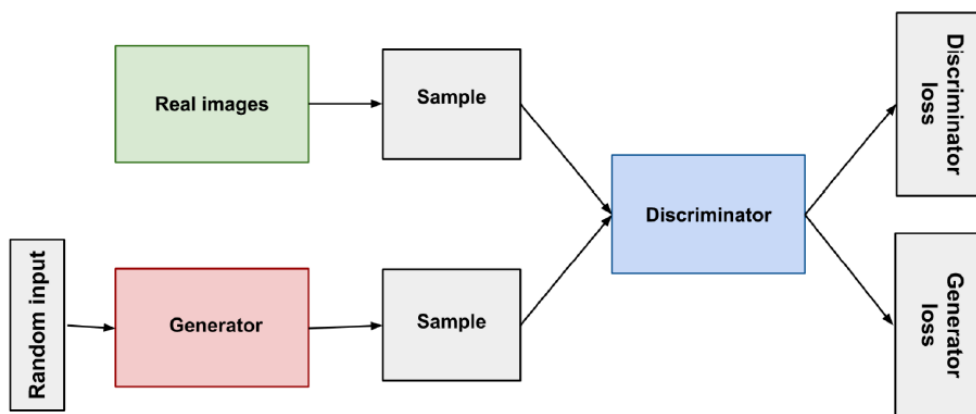


图 4 生成对抗网络结构图

Fig.4 The struct of Generative Adversarial Networks

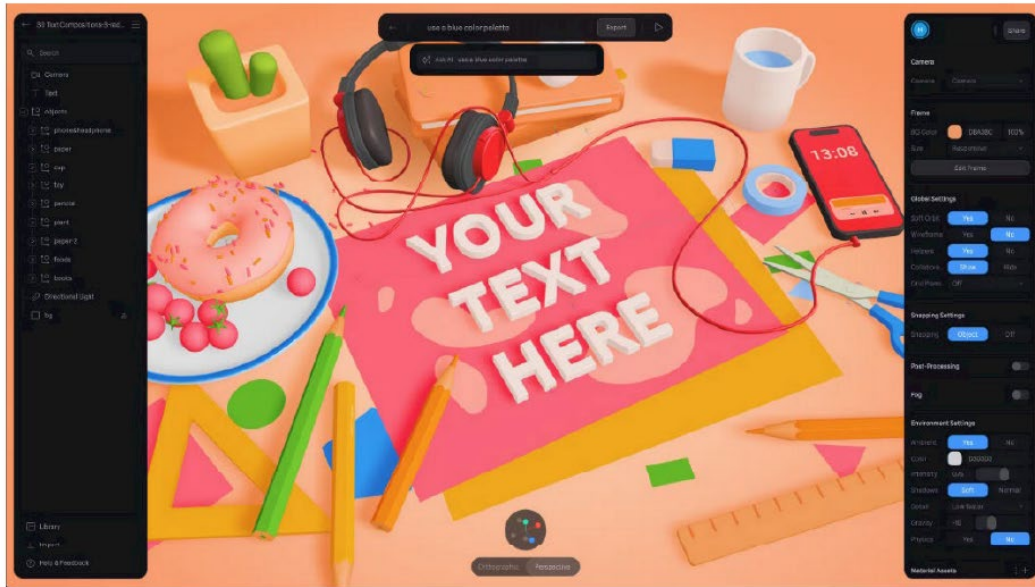


图 5 Spline AI 操作界面

Fig.5 Spline AI Operation Interface

型数据集的特征和规律,生成新的三维模型。与传统的三维建模工具相比,Spline AI 可以更快、更精准地生成高质量的三维模型。除了生成三维模型外,Spline AI 还提供了丰富的编辑工具,如平滑、拉伸、旋转、缩放等,可以帮助用户对生成的模型进行微调和优化。此外,Spline AI 还支持导入和导出多种三维模型格式,如 OBJ、FBX、GLTF 等,方便用户在不同的应用场景中使用生成的三维模型。

4.3 AIGC 模型生成技术展望

虽然 AIGC 技术在三维模型生成领域已经展示出很大的优越性和潜力,但是仍然存在一些问题和挑战需要进一步研究和解决。

首先,数据集的构建仍然是一个困难和耗费成本的问题。构建高质量的三

维数据集需要耗费大量的时间和资源,且数据集的规模、多样性和真实性等方面需要考虑。因此,如何更好地构建高质量的三维数据集是一个亟待解决的问题。

其次,AIGC 技术在三维模型生成中仍然存在着模型的不稳定性和过拟合问题。在训练过程中,生成模型容易出现各类等问题导致生成的三维模型质量下降。因此如何提高生成模型的稳定性和泛化能力是未来的一个重要问题。

人工智能模型在生成内容时常常存在“风格单调”的问题,即内容通常缺乏多样性和创新性,呈现出相对单调的风格。在三维模型生成领域内,现阶段市面上的各类新型 AI 工具都具有这样的问题,如 Spline AI 生成的三维模型具有饱和、圆润的卡通风格,而 Opus AI 生成的模型具有写实风的特征。

AIGC 技术在三维模型生成领域还

存在着一些瓶颈。例如，如何更好地处理三维模型中的几何形状和纹理信息，如何实现模型的自动化选择和优化，如何快速生成高质量的三维动态模型等，都是需要进一步研究和解决的问题。

未来，随着硬件设备和算法技术的不断提升和发展，AIGC 技术在三维模型生成领域的应用将会有更广阔的前景和更多的研究热点。例如，如何结合增强学习算法和迁移学习算法等技术，进一步提高模型的自适应性和泛化能力；如何实现对三维模型的精细控制和交互式编辑等功能，以满足更多应用场景的需求；如何实现快速生成高质量的三维动态模型等，以扩展三维模型生成的应用范围和实现更高效的三维模型生成等。

5. 三维场景生成

三维场景生成最关键的部分是场景理解与场景合成。前者需要考虑大量的场景信息，选出合理的对象类别，后者

则需要适当的将选出的对象布置在场景中。三维场景可以说是虚拟现实的重要基础，一个好的三维场景能够极大的提升虚拟现实技术带给使用者的体验。因此，三维场景的生成技术在不断的推陈出新，效果逐渐提升，但其本身的复杂性也带来了许多的挑战，如仿真场景的真实性、虚拟场景的合理性以及场景本身的美化程度，都需要大量的人力、物力投入。目前，Chat GPT 及其带来的 AIGC 技术给场景理解和场景合成都带来了新的发展。

5.1 传统三维场景技术

传统的三维场景生成技术中，人机交互方法占据了主要的地位，场景理解主要由开发者完成，借助 3D 建模软件或是一些专用软件进行自动或是半自动的场景合成工作。

传统技术在三维场景生成方面有着广泛的应用。美国伊利诺伊大学的 Fard、Mani 等人将 BIM 技术和摄影影

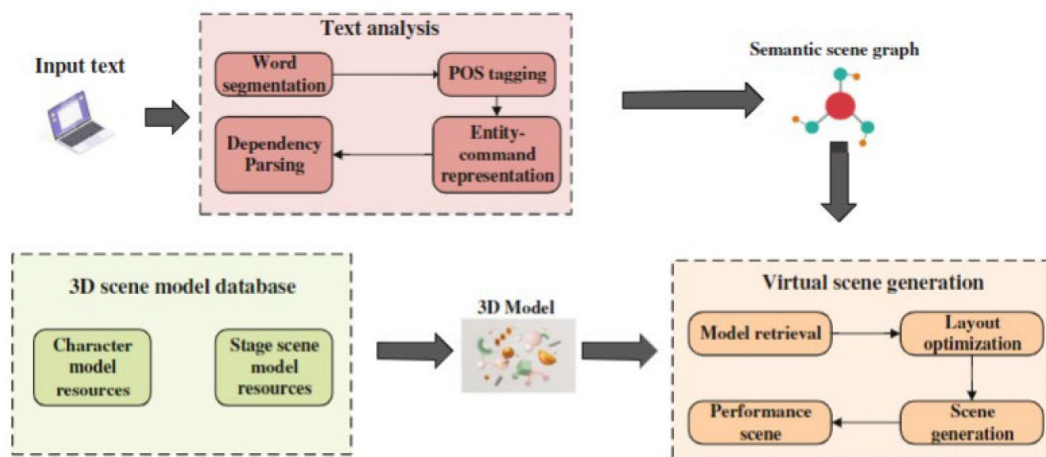


图 6 由文本生成戏曲三维场景过程

Fig.6 Overview of text to Chinese opera 3D scene generation

像技术相结合, 利用软件处理数据资料以完成 3D 施工的模拟效果^[10]; 基于三维软件 sketchup 可建立小型建筑场景, 实现建筑场景的三维可视化表达; 许多大学, 也采用全景技术搭建校园场景, 方便新生及其他社会人士, 在任何地点通过虚拟现实技术了解校园场景。

我们可以看到, 开发者需要承担许多重复、繁琐的场景搭建工作, 且场景搭建的效果与开发者对场景的理解有着相当大的关联, 有着极大的潜在风险。与此同时, 通过这些方式生成的三维场景大都是静态的, 与用户存在的交互很少且体验效果较为机械。目前的虚拟现实市场对于更优的交互、更真实、动态变化的场景有着较高的需求。而动态化的三维场景搭建, 需要新的技术或方法来提供更好的解决方案。

5.2 基于语义的戏曲表演生成探索

Chat GPT 表现出的强大学习能力、语义理解能力、上下文记忆对话能力带来极大震撼的同时, 也让我们看到了其所带来的 AIGC 技术的强大生产能力。三维场景生成中的场景理解可由 AIGC 完成, 且同步自动进行后续场景合成等, 完成三维场景生成工作。AIGC 技术给三维场景合成带来了新的发展动力。

来自郑州轻工业大学、英国伯恩茅斯大学的梁辉等人, 将 AIGC 技术应用于三维场景生成, 提出了一种基于语义

的中国戏曲表演生成方法^[11]。该方法可依据输入的描述性脚本自动生成戏曲表演的三维场景, 首先通过分析场景的描述脚本了解戏曲虚拟场景的构成要素, 然后利用先验概率学习场景模型中的模型放置规则, 最后生成出与脚本对应的戏曲表演的数字场景。该方法被证明可以生成自然的可接受的歌剧数字表演场景。

该方法由三部分组成: 由戏曲数字资产三维模型本体库支撑的数据库; 由用户输入脚本理解出具有空间关系的语义场景地图; 由语义场景地图从数据库中提取模型并进行场景搭建。流程图图 6 所示。

场景理解方面, 该方法的输入脚本处理分为两步完成。第一步是利用隐马尔可夫模型将输入的描述脚本划分为独立的虚拟场景元素, 然后进行顺序标注, 实现元素分割。第二步是进行语义统计及独立元素分类, 将输入的文本信息理解成空间语义信息, 如图 7:



图 7 空间语义解析示例

Fig.7 Example of spatial semantic parsing

场景合成方面, 该方法对数据库中的每个模型都添加了名称、属性、向前方向、大小、旋转和质心坐标等语义标记, 将语义标签与数据库中的模型相关联。在此基础上, 就可基于前述的空间语义信息生成具体的三维模型。



图8 场景合成示例

Fig.8 Example of scene composition

通过此方法，我们可以看到，良好的场景理解可以为自动的场景合成提供优良的基础。

5.3 AIGC 场景生成技术展望

可以期待，如 Chat GPT 的大型语言模型的语义理解能力会不断增强，可以准确、高效的根据三维场景的需求完成场景理解工作，深度学习技术的发展会让场景合成不再需要大量的人力投入。AIGC 技术将会使三维场景生成工作高

度自动化，用户只需输入他的需求，便可得到与之对应的精细的三维场景。开发者的工作将由具体的场景搭建，转变为建立三维场景自动搭建的框架。

但同时，自动化的生成工作也势必带来不可预知的问题。一方面，由于生成内容完全依赖于用户输入，就有可能导致结果同质化，缺乏多样性，也可能导致生成场景的效果无法达到用户需求，缺乏可控性。另一方面，用户的输入决定输出，也意味着自动生成场景的结果在很大程度上是不可控的，当下已经出现利用 AI 技术生成假视频的现象，如何将自动生成的结果控制在社会道德乃至法律范围之内将是一项必要工作。

总之，Chat GPT 及其带来的 AIGC 技术给三维场景生成技术的发展带来了新的动力，自动化的三维场景生成已成为一种发展趋势。未来，如何实现更精

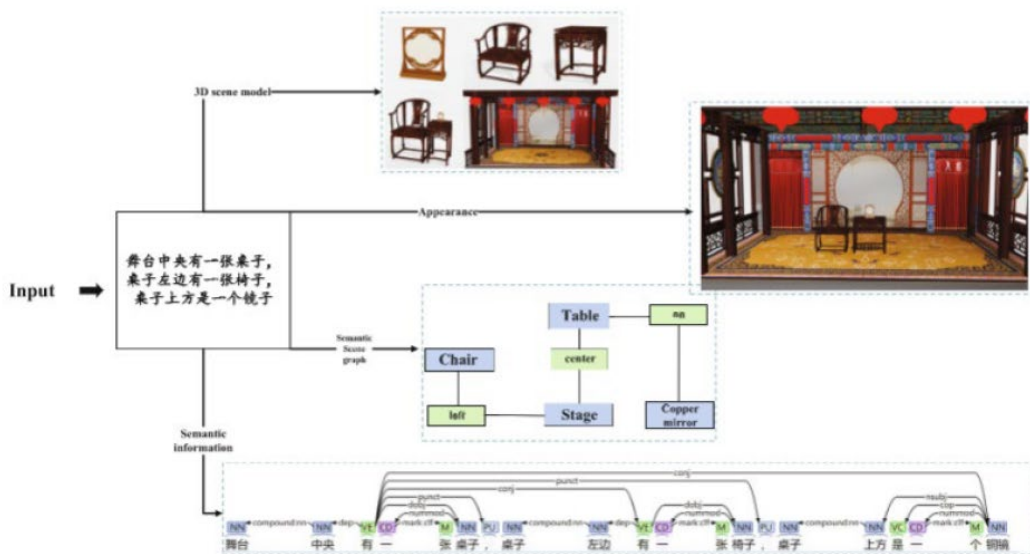


图9 舞台场景合成过程

Fig.9 Stage scene synthesis process

细的场景理解，如何利用深度学习技术实现更好的动态化场景搭建，以及如何进行有效的监管，将有很大机会成为热点话题。

6. 虚拟对话生成

作为在近年取得了重大突破并成为研究热点的技术，AIGC 技术很有可能在不久的将来对虚拟现实技术产生重大影响，使得虚拟世界对于用户来说更加具备沉浸感，并且大大改善人机交互体验。近期以来，Chat GPT 的优秀表现或许标志着 AIGC 技术已经能达到应用于虚拟世界的对话生成。

通过 AIGC 技术，虚拟世界的开发者们可以创造更为真实、互动性更强的虚拟人物。传统的 VR 体验中，用户与虚拟角色之间的对话是脚本化的，这意味着用户与虚拟角色之间的交互总的来

说由程序预设直接决定，限制了用户与虚拟人物的互动范围。AIGC 技术的发展，使得虚拟角色能够通过 NLU

（Natural Language Understanding，自然语言理解）技术和 NLP（Natural Language Processing）技术，扩大用户的输入类型并且做出动态反应。例如，通过 NLU 和各种人工智能技术，虚拟角色可以理解和回应用户的对话输入并且作出反应，极大改善用户与虚拟世界之间的交互形式。

6.1 AIGC 应用于虚拟世界的探索

NLU 和 NLP 的快速发展使得机器在处理自然语言方面获得了接近于人类的水平，也让人们看到了在虚拟世界中创造更为真实的虚拟角色的希望。目前，已经有一些研究人员着手于将



图 10 虚拟村庄社区截图

Fig.10 Screenshot of Virtual Village Community

AIGC 应用于虚拟世界，通过创建大量的 AI 角色并令他们交互来构造一个 AI 世界。

受到游戏《模拟人生 (The Sims)》的启发，为了创建一个更具有真实感的人工 AI 社区，研究人员创建了一些 AI 角色，为它们制定行为规则，引入 AIGC 功能并定义它们之间的交互行为，尝试创建一个虚拟村庄。^[12]

研究人员在此篇研究中指出，由于人类行为所涉及到的空间十分巨大且复杂，尽管目前的 LLM 已经能够在单个时间点以十分优异的表现模拟人类行为，但目前的 AIGC 难以在一个长期维持的虚拟社区中保持合理的行为序列。这是因为随着时间的推移，人类的行为会受到越来越多的相关事件和互动的影响，并且通过反思这些事件和进行概括，以进行更高层级的推论。因此，相较于现有的 AIGC 技术，虚拟角色需要拥有一定时间的“记忆”，并通过这些记忆影响自身产生答案的结果。

基于以上理论，研究人员将虚拟人物的代理结构分为三大部分：记忆流、反射和计划。在这之中，记忆流模块使用自然语言来长期记录一个全面的经验清单，并且通过检索的方式结合相关性、时间的接近性和重要性来显示这些经验需要通知角色的即时行为。反射模块负责在时间经过的同时，将记忆模块中的经验清单选择并且生成更高层次的推理，使得虚拟角色能够得出结论、了解自己和他人。计划模块将反射模块所得到的结论保存为更高级的行动计划，

并且递归地转换为自身行动和做出反应的详细行为。

除了研究人员所进行的工作之外，一些网络技术人员也在尝试在数字世界构建 AI 社区，例如近期在互联网上出现的只允许 AI 进行交互的虚拟社区 Chirper。Chirper 作为 AI 社交网络，只允许人类在此社交网络上创建 AI，使得这些角色能够参与到虚拟社区的讨论当中。

在这些尝试当中，我们惊奇地发现了与传统的、模板化、脚本化的 AI 社区不同的新型 AI 社区，并且认为这是 AIGC 技术应用于虚拟现实世界的前瞻。

6.2 AIGC 对话生成技术展望

通过前文的讨论，我们已经了解 AIGC 技术在 NLP 领域和 NLU 领域取得的巨大突破和近期研究者将其应用于创建人工 AI 社区的尝试。在这一节我们将会针对这一技术目前面临的一些困难作探讨，并且展望技术在未来可能的发展。

目前，AIGC 应用离良好地应用于虚拟世界依然存在以下技术壁垒：

(1) 记忆的保持和根据经验所做出高层次推断的能力。在前文的虚拟村庄的构建中，研究人员指出目前 LLM 模型所能执行的语言生成任务只能针对时间跨度极小的紧急任务中获得良好的效果。尽管虚拟村庄的创造者们给出了让 LLM 维持一段时间的记忆并且形成反

射和计划的方案, 但此方案依然存在不小的失败率, 此方案仍需要进一步的研究以获得更好的表现效果。

(2) 更为丰富和强大的共情能力和情绪化表达。研究人员指出, 人工智能系统在驾驭复杂的人类情感方面仍然面临困难, 由于这类任务的开放性, 人类与人工智能之间的情感交流工作仍然具有非常大的困难, 例如包含情感宣泄的对话^[13]。研究者正尝试改善模型, 提高会话之间的共情率以使得人工智能能够在与人类的情感沟通之中取得更好的表现。

从另一个角度来说, AIGC 的发展, 尤其是 GPT 模型在近年来取得的突破, Chat GPT 和 GPT-4 进行了大规模的预训练, 从互联网中获取了海量的知识, 并且能够从人类的反馈中进行强化学习, 进一步增强了 GPT 模型的适应性和性能, 使得其在处理自然语言的方面获得了很高的效率和强大的能力。在未来, 随着 GPT 的应用面走出文本任务, 扩展到教育、历史、数学、物理学、编码等各种领域, AIGC 将能够创造出更多更丰富、更具有特点、更复杂的角色; 以及 AI 对于人类情感理解能力的进一步增强。从而使得虚拟角色能够更千变万化, 更加具有真实感, 更能够使用户沉浸于人类与虚拟角色之间的交互之中。很有可能在不久的将来, 人类将会拥有创造具有情感、知识丰富、性格特征鲜明的虚拟人的能力, 使得人类与虚拟人之间的交互如同人类与人类之间的交互。而人类与虚拟角色之间的

关系也将随着相关道德和法律完善的同时进一步走向合理。虚拟角色既可以成为人类的老师, 也可以成为在虚拟世界中与人类密切接触, 相互交流情感的朋友。

7. 总结与展望

AIGC 具备强大的生产力、个性化的生产过程和优质可体验的生产结果, 是其关键优势。它能实现全天候、快速的内容生产, 并具有自主性和多样性。与传统模式不同, AIGC 通过智能个性化生产避免了同质化问题。它提供高质量的内容, 通过算法优化和技术支持实现更好的视觉效果, 并与 VR、AR 技术结合提供三维沉浸式体验。总之, AIGC 在内容生成领域具备重要竞争优势。

基于 AIGC 技术的模型生成技术通过训练深度学习模型, 实现了自动建模的革新。这种技术具有高效、快速和可扩展等显著优势, 可以有效解决了传统三维模型生成技术所面临的问题。相较于传统的手动建模方式, 该技术在效率上有了明显提升, 同时也能克服数据获取和处理的困难, 以及几何形状和纹理信息受限的限制。基于 AI 的三维模型生成技术正不断创新和发展, 在生成复杂的几何形状和纹理信息方面取得了令人瞩目的进展。

基于 AIGC 技术, 虚拟世界的开发者们可以创造更为真实、互动性更强的虚拟人物。传统的 VR 体验中, 用户与

虚拟角色之间的对话是脚本化的,这意味着用户与虚拟角色之间的交互总的来说由程序预设直接决定,限制了用户与虚拟人物的互动范围。AIGC 技术的发展,使得虚拟角色能够通过 NLU 技术和 NLP 技术,扩大用户的输入类型并且做出动态反应。例如,通过 NLU 和各种人工智能技术,虚拟角色可以理解和回应用户的对话输入并且作出反应,极大改善用户与虚拟世界之间的交互形式。

未来,在 VR&AIGC 领域,还有不少亟待解决的问题与广阔的发展前景:

(1) 如何结合增强学习算法和迁移学习算法等技术,进一步提高模型的自适应性和泛化能力;如何实现对三维模型的精细控制和交互式编辑等功能,以满足更多应用场景的需求;如何实现快速生成高质量的三维动态模型等,以扩展三维模型生成的应用范围和实现更高效的三维模型生成等。

(2) 记忆的保持和根据经验所做出高层次推断的能力。在前文的虚拟村庄的构建中,研究人员指出目前 LLM 模型所能执行的语言生成任务只能针对时间跨度极小的紧急任务中获得良好的效果。尽管虚拟村庄的创造者们给出了让 LLM 维持一段时间的记忆并且形成反射和计划的方案,但此方案依然存在不小的失败率,此方案仍需要进一步的研究以获得更好的表现效果。

(3) 更为丰富和强大的共情能力和情绪化表达。研究人员指出,人工智能系统在驾驭复杂的人类情感方面仍然面临

困难,由于这类任务的开放性,人类与人工智能之间的情感交流工作仍然具有非常大的困难,例如包含情感宣泄的对话。不少研究者正尝试改善模型,提高会话之间的共情率以使得人工智能能够在与人类的情感沟通之中取得更好的表现。

(4) 目前 VR 设备仅仅捕捉用户简单的输入信息,并未捕获用户的眼神、面部表情和心率等生理指标。未来 AIGC 还可以通过分析这些参数,实现更加精准的情感识别。这使得虚拟现实体验更加个性化和逼真,进一步提升沉浸感和参与度。

8. 参考文献

- [1] 黄冠, 曾靖盛. 虚拟现实技术的研究现状、热点与趋势[J]. 中国教育信息化, 2022, 28(10): 49-57.
- [2] 欧阳春雪. 基于深度学习的多模态 AIGC 动画探究[J]. 现代电影技术, 2023, No.534(01): 41-47.
- [3] Sakib Shahriar, Kadhim Hayawi. Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations[EB/OL]. arXiv, 2023.3.2. <https://arxiv.org/abs/2302.13817>
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. Arxiv, 2019.5.24. arXiv:1810.04805v2
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. Arxiv, 2019.5.24. arXiv:1810.04805v2
- [6] 郭全中, 袁柏林. AIGC 与 WEB3.0 有机融合: 元宇宙内容生产的新范式[J]. 南方传媒研究, 2023, No.100(01): 36-47.
- [7] Li C, Zhang C. When ChatGPT for Computer Vision Will Come? From 2D to 3D[J]. arXiv preprint arXiv:2305.06133, 2023.
- [8] Achlioptas P, Diamanti O, Mitliagkas I, et al. Learning representations and generative models for 3d point clouds[C]//International conference on machine learning. PMLR, 2018: 40-49.
- [9] Li C, Zhang C, Waghware A, et al. Generative AI meets 3D: A Survey on Text-to-3D in AIGC Era[J]. arXiv preprint arXiv:2305.06131, 2023.
- [10] 鲁鑫鑫. 虚拟三维场景的搭建与可视化动态展示[D]. 山东大学软件工程, 2021.
- [11] Liang H, Dong X, Liu X, et al. A semantic-driven generation of 3D Chinese opera performance scenes[J]. Computer animation and virtual worlds, 2022(3/4): 33.
- [12] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior[EB/OL]. arXiv, 2023.4.7. arXiv:2304.03442v1
- [13] Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, Tim Althoff. Human-AI Collaboration enables more empathic conversations in text-based peer-to-peer mental health support[J]. Nature machine intelligence, 2023, 5, 46-57