

智能艺术工厂：基于平行艺术智能体的创作自动化

郭超¹, 戴星原¹, 廖玮琰², 祁珏瑜², 鲁越³, 倪清桦⁴, WANG Lin⁵, 王飞跃¹

(1. 中国科学院自动化研究所, 北京 100190;

2. 北京科技大学, 北京 100083;

3. 山东大学控制科学与工程学院, 山东 济南 250061;

4. 澳门科技大学, 澳门 999078;

5. 加州大学伯克利分校劳伦斯国家实验室, 美国 伯克利 94720)

摘要: 近年来, 多模态内容生成模型迅速发展, 极大地提高了创意工作者的创作效率。然而, 现有方法仍然需要领域专业知识, 无法实现创作全流程的自动化, 难以广泛服务于普通人群。因此, 提出智能艺术工厂的理念和架构, 基于大模型和人工智能智能体技术, 构建智能艺术工厂的工作流程、基本功能和实现方法, 建立面向创意创作的智能体员工和数字人团队, 通过生物人、数字人和机器人的跨物理、信息、社会空间协作, 实现创作流程自动化和创作知识自动化。智能艺术工厂将进一步解放大众的创造力。

关键词: 智能艺术工厂; 平行艺术; 基础模型; AI 智能体; 人机协作创作

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202421

Intelligent art factory: achieving creative automation through parallel art agents

GUO Chao¹, DAI Xingyuan¹, LIAO Weilong², QI Jueyu², LU Yue³, NI Qinghua⁴, WANG Lin⁵, WANG Fei-Yue¹

1. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

2. University of Science and Technology Beijing, Beijing 100083, China

3. School of Control Science and Engineering, Shandong University, Jinan 250061, China

4. Macau University of Science and Technology, Macao 999078, China

5. Lawrence Berkeley National Laboratory, University of California, Berkeley, Berkeley 94720, USA

Abstract: In recent years, multimodal content generation models have rapidly evolved, significantly improving the efficiency of creative workers. However, existing methods still require expertise and cannot realize the automation of the whole creation process, making it difficult to widely serve the public. This paper proposes the concept and architecture of intelligent art factory, based on advanced foundation models and AI agent systems. Creative production oriented agents and digital teams, in conjunction with the basic workflow, functions, and methods of the intelligent art factory are established. The creative process automation and creative knowledge automation are achieved through collaboration among biological humans, digital humans, and robots in cyber-physical-social spaces. The intelligent art factory seeks to further liberate the creativity of the public.

Key words: intelligent art factory, parallel art, foundation model, AI agent, human-machine co-creation

收稿日期: 2024-04-06; 修回日期: 2024-05-22

通信作者: 王飞跃, feiyue.wang@ia.ac.cn

基金项目: 国家自然科学基金项目 (No.62306319)

Foundation Item: The National Natural Science Foundation of China(No.62306319)

0 引言

近年来,随着生成模型等人工智能方法的快速发展,Stable Diffusion、Midjourney、Sora等现象级的技术应用^[1-2],覆盖了图像、视频等不同模态的内容生成,给人类的创作性工作带来了极大的便利,大幅降低了大众进行艺术创作的门槛。这些生成模型可能在特定内容的生成上达到媲美人类设计师的水平,但面对复杂的艺术创作流程和内容时,需要精通专业领域知识并基于提示工程进行反复尝试,通过大量工具和模型的设计与拼接才能完成创作的整个 workflow,无法实现创作全流程的自动化以及以用户为中心的创作知识自动化,难以广泛服务于普通人群。此外,目前的艺术内容生成仅限于信息空间内的数字媒体,无法与人在物理空间内协同实现实物创作。因此,如何利用AI技术进行人机物协同,实现艺术创作流程自动化和创作知识自动演化,具有重要意义。构建跨信息—物理—社会空间、可计算、可信可靠的艺术辅助创作系统,将极大地扩展艺术创作的边界,以机器的训练与迁移模拟创作的练习与演化,以机器的计算和执行提升人类的创作效率,解放每一个人的创造力。

在以AlphaGO为代表的算法智能之后,相继出现了以ChatGPT为代表的语言智能^[3]和以Sora为代表的想象智能^[4]。这些语言模型、多模态大模型展现出强大的跨模态感知能力、建模表征能力、逻辑推理能力与表达生成能力^[5-6],并且能够通过思维链(chain of thought, CoT)等机制进行自我引导和优化。同时,大模型技术被应用于实际物理信息系统的决策任务,例如利用智能机器人完成炒菜、打扫卫生等复杂居家任务。得益于这些通用人工智能方法的发展,人工智能智能体(AI agent)能够以具备特定能力的大模型为基础,通过任务规划、工具调用、记忆反思等机制,组织各类模型和工具应对复杂的工作流程^[7-8],并通过自我迭代、演化提升整体性能^[9]。因此,基于大模型的人工智能智能体为艺术工厂的构建提供了可行的路径。

一方面,智能体系统通过调用大模型、机器人等已有的基本方法和工具,跨物理、信息、社会空间实现人机交互以及整个创作流程的自动化^[10]。另一方面,数字人团队及其各级智能体之间通过交互展开大规模艺术创作计算实验以及对创作流程的

探索,通过迭代优化实现创作技能和知识的自动演化及提升,使得创作知识自动化成为可能。在此基础上,形成以人为中心的智能艺术工厂。继解放生产力之后,智能艺术工厂将解放创造力。

本文基于平行艺术架构和人工智能智能体系统方法构建智能艺术工厂,建立与实际艺术创作过程相对应的虚拟人工系统,设计生物人、数字人、机器人协同的艺术工厂底层架构,构建艺术创作智能体及其团队,并阐述艺术工厂的基本工作流和实现方法。

本文的组织结构如下:第1节综述人工智能智能体和人机协同创作的相关工作;第2节介绍平行艺术架构;第3节阐述艺术工厂的底层架构、基本构成、工作流程和实现方法;第4节总结全文。

1 相关工作

1.1 AI智能体

得益于大语言模型在文本生成、知识获取、多轮对话、逻辑推理等任务上的卓越能力^[6],AI智能体在视频剪辑^[11]、游戏^[12]等领域的各类复杂任务中取得了突破性进展^[13-14]。Yao等^[15]提出ReAct框架,通过生成交错的推理序列和行动,优化智能体的规划、推理和行动过程,提高了智能体的整体性能和可控性。Wang等^[16]提出融合分析模块、记忆模块、规划模块和行动模块的大语言模型人工智能代理统一框架,采用语言模型进行智能体的控制,并通过多模态感知和工具利用等策略扩展其感知和动作空间。Wei等^[17]通过思维链和问题分解等机制进一步提升智能体的推理和规划能力。

然而,单一智能体缺乏合作,难以通过交互协作与信息共享获取足够的知识和经验^[18],导致其无法被应用于复杂的场景。鉴于此,研究者提出采用多智能体系统进行互补性交互^[19-20]。智能体评估其他智能体的需求与能力,并主动寻求合作和信息共享,以调整其行为策略,从而满足任务需求。此外,部分研究探讨如何通过智能体与人之间的交互提升自身工作效率^[21],人类向智能体提供必要的引导提示,并基于定性或定量评估提供反馈,进一步优化智能体的行为决策。

针对艺术创作过程需要多功能配合和多轮迭代优化的问题,本文利用特定功能的智能体协同完成复杂的创作流程,并通过智能体间的交互演化和迭代优化来提升各智能体和艺术工厂的整体创

造能力。

1.2 人机协同艺术创作

现有的人机协同创作主要体现为提供创作建议、学习人类创作行为并辅助人类创作、人机共同完成创作过程等。

Parikh等^[22]设计了一种机器投票机制用于多人协作创作,提升了多人交替简笔画的最终视觉效果。一些研究基于用户已有的绘画内容,考虑绘画的局部和整体相似度,在绘画数据库中检索专家作品,并实时推荐给用户作为创作参考,为用户提供语言或视觉引导^[23-25]。Karimi等^[26]为用户推荐同类的绘画选项,为用户提供更多的创作灵感。

Co-GAIL^[27]提出了一种从人与人的协作演示中学习人与机器人的协作策略的方法。该方法通过互动学习同时优化人类策略和机器人策略:人类策略学习从演示中产生的多样化、可信的协作行为;机器人策略通过预估人类合作者的潜在策略来学习如何进行协助。Liu等^[28]研发了面向盲人的绘画导航系统,通过视觉感知画面和画者状态,并通过语音引导画者持笔绘画。Suh等^[29]研发的CodeToon工具可以将编程代码以漫画的形式呈现,帮助用户更好地理解编程概念。此外,Chakrabarty等^[30]探索机器如何为人类写作提供规划、翻译、审阅等辅助功能。

上述人机协作方面的研究侧重于通过AI为人类提供一定的创作灵感和创作辅助服务,忽略了人和机器在知识和技能上的优势互补,且难以应对开放性创作任务。在智能艺术工厂中,本文基于多智能体系统对机器模仿人类进行创作和机器协同创作演化进行探索。

2 平行艺术: 智能艺术工厂的基础

2.1 平行艺术的基本结构

为了提升机器的创造能力,在艺术创作中实现双向的人机协作,本文基于平行系统和ACP方法(人工系统,计算实验,平行执行)^[31-32]提出了平行艺术架构^[33]。其核心在于针对实际艺术创作过程建立与之对应的虚拟人工系统,通过人工系统中的计算实验^[34]对实际艺术创作过程进行探索、分析和评估,扩大了艺术创作的探索空间并提高了探索效率。平行艺术通过人工系统(虚拟)与实际系统(现实)之间的平行执行、虚实互动、迭代优化引导实际创作过程,以人机混合协作的方式提升机

器的创造能力^[10]。

如图1所示,平行艺术架构包括描述、预测、引导和评估4个关键阶段。1)描述阶段,采集人类创作过程的真实数据,建立反映实际创作过程的虚拟人工系统。2)预测阶段,在构建的人工系统中进行大规模计算实验,广泛尝试不同的创作方案,并向实际系统推荐优化的创作策略。艺术创作的计算实验可以突破实际创作中的物理、时间和成本限制,从而扩大其探索空间并提高创作效率。3)引导阶段,在虚拟人工系统的指导下,执行经过计算实验探索和验证的创作策略并不断对其进行迭代更新,以提高实际创作效果。4)评估阶段,人类艺术家对计算实验推荐的创作方案和实际创作结果进行评估,为人工系统和创作策略的迭代优化提供人类反馈。人工系统和实际系统并行执行、迭代优化,这意味着人工系统根据对实际系统的观察进行自我更新,而实际系统执行人工系统推荐方案进行实际创作。

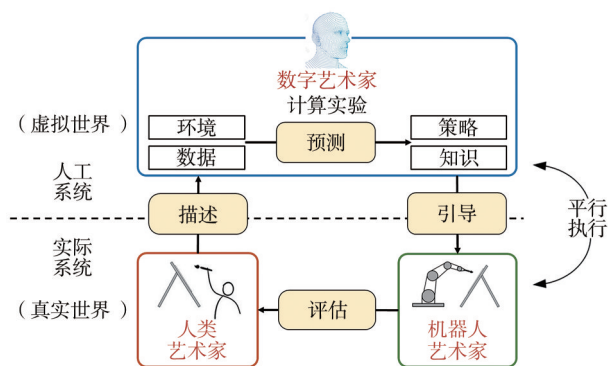


图1 平行艺术架构^[10]

目前,平行艺术初步应用于人机协同绘画^[35]、戏剧创作^[36]等领域^[37-38]。本文基于平行艺术架构研发了绘画机器人系统,实现了机器对人类绘画创作过程的学习、复现,其绘画技能达到了人类水平^[39],并在此基础上构建了艺术大模型,实现了人机协同创作^[40]。

2.2 平行艺术中的3类角色

在平行艺术架构中,人类艺术家、数字艺术家和机器人艺术家3类角色进行跨物理、信息、社会空间的混合作业。数字艺术家是人类艺术家在信息空间的“分身”,拥有人类艺术家的长期创作数据和定制化知识模型。它在人工系统中代表人类艺术家进行大规模创作实验,实现对创作技能的探索与

演化。机器人艺术家是人类艺术家在物理空间的“分身”，替代人类艺术家执行物理创作过程，将人类从烦琐的体力劳动中解放出来，从而将精力集中在需要深度创造力的部分。人类艺术家是艺术创作的主体，与数字艺术家、机器人艺术家形成双向协同。一方面，人类艺术家为机器（数字和机器人艺术家）提供创作示范，从而提升机器创作水平。另一方面，人类艺术家对数字和机器人艺术家的创作过程和结果进行评估，通过人类反馈改进其创作技能和创作风格，将数字和机器人艺术家演变为人类艺术家的个性化定制“助手”。人类和机器通过协同合作，螺旋式地提升各自的艺术创作能力，最终演化形成的平行艺术系统具备与人类相近的创作技能，进一步辅助人类进行艺术创作。

平行艺术方法为智能艺术工厂的实现奠定了基础。人类艺术家主要进行技能示教、价值引导、效果评估，数字艺术家主要进行需求交互、知识积累、技能学习和基于计算实验的大规模创作探索。机器人艺术家主要在物理空间下协助人类进行实体创作，减轻人类的工作负担。人类艺术家、数字艺术家和机器人艺术家在虚拟人工系统和实际系统中的协同配合和迭代优化，有助于实现智能艺术工厂的创作流程自动化和以用户为中心的创作知识自动化。

3 智能艺术工厂

在平行艺术架构的指导下，本文以多模态大模型为基础模型，融合艺术创作模型库和工具，组建艺术创作的智能体和数字人团队，通过生物人、机器人、数字人之间跨物理、信息、社会系统的协作

来实现创作流程自动化和创作知识自动化，提供辅助创作服务，释放大众的想象力和创造力^[10]。

本节将阐述人机物协同的智能艺术工厂的底层架构，在此基础上构建艺术创作智能体和数字人团队，然后介绍智能艺术工厂的工作流与功能，最后阐述智能艺术工厂的实现方法。

3.1 智能艺术工厂的底层架构

在艺术工厂中，生物人、机器人、数字人3类角色互相协作。其中，数字人是艺术工厂的执行者和维护者，生物人主要面向数字人进行需求提示、方案批准、技能示教、状态监控、价值引导、效果评估等，机器人作为数字人可调用的物理工具，用于执行具体的物理交互任务。艺术工厂存在3种运行模式：自主模式（autonomous mode, AM）、平行模式（parallel mode, PM）、专家/应急模式（expert/emergency mode, EM）。艺术工厂在大部分情况下处于AM模式，数字人根据用户需求自动调用工厂资源完成艺术创作全过程。而当数字人遇到无法解决的难题时，工厂或其中某些子系统切换到平行模式，由生物人远程操纵和指导数字人完成困难任务，并将操作过程补充到创作经验库中。而专家模式用于工厂构建初期对新智能体的教学，应急模式面向需要人类亲临现场的复杂任务。智能艺术工厂的底层架构如图2所示。

数字人由基于不同模型、算法、工具构建的智能体组成，艺术工厂包含负责创作任务规划的规划智能体、负责具体任务执行的执行智能体、负责创作测试和评估的评估智能体、负责人机交互的交互智能体等。针对某一用户提出的创作需求，交互智能体通过与人的多轮迭代交互明确任务目标，并由规划智能体根据已有的创作知识针对特定目标进行

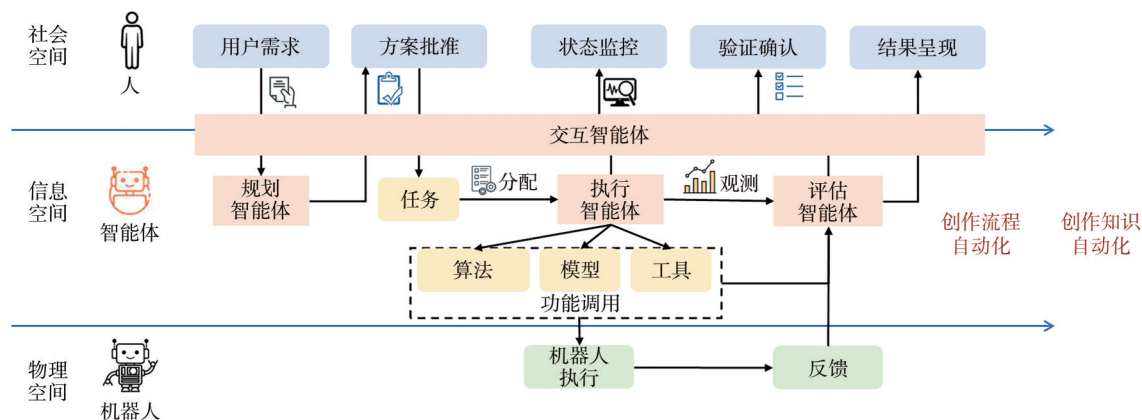


图2 智能艺术工厂的底层架构

创作任务规划，创建任务流，经过人批准同意后，将各级创作任务下发给不同的数字人团队和执行智能体。执行智能体利用角色调用算法、模型和工具完成分配的局部子任务，包括指令控制机器人实体完成物理空间内的创作并提供状态反馈。评估智能体根据执行过程中监测的各智能体和机器人的任务状态和最终结果进行性能统计和评估，从而改进技能以及将必要的创作流程反馈给人进行调整。与此同时，交互智能体实时监控艺术工厂的运行状态，将关键环节交给给人进行验证确认，并呈现创作过程和结果。

智能体可以进行自由组合调度，以任务为导向动态形成团队，调用不同功能的模型、算法和工具完成分配的子任务，通过各级智能体和任务导向的团队之间的配合协作，实现整个创作流程的自动化。

智能体的训练和学习通过数字艺术学校（digital art school, DAS）实现：首先，根据实际的艺术创作应用建模创作技能场景，构建艺术训练数据和交互练习环境，设计相关的创作工具和算法，进一步训练创作技能模型；其次，根据智能体对应的功能角色，建立其与特定工具库、算法库、模型库的关联关系，并通过虚拟人工创作系统中的练习环境进行技能调用训练，实现对各级不同功能智能体的训练。

3.2 智能艺术工厂的构成

基于上述智能体构建具备不同艺术创作功能的数字人团队，主要包括任务规划团队、创意生成团

队、创作执行团队、审美鉴赏团队、数据处理团队、人机交互团队等，通过各团队间的协作配合以及团队内智能体对各类基础模型和工具的调用，完成艺术创作的“流水线”。智能艺术工厂的构成如图3所示。

（1）任务规划团队

任务规划团队根据用户需求进行任务分析和分解，由人机交互团队与用户确认任务规划方案，并进行迭代优化，最终创建任务流，将拆分的任务流分配给各团队和对应功能的智能体。该团队需要对艺术创作流程有充分的认知和逻辑规划能力，主要基于检索增强生成（retrieval-augmented generation, RAG）方法增强的语言大模型实现。

（2）数据处理团队

数据处理团队采集海量公开的艺术史、艺术评论、艺术作品等多源异构数据并进行清洗整理，融合用户个性化的隐私数据，通过编码器编码构建RAG向量数据库，并在不同用户之间形成分布式联邦艺术知识库^[41]。在艺术工厂运行过程中，根据服务人群进行个性化的素材检索，对其他团队进行数据和知识增强，并利用实时创作的数据与知识更新数据库，为创作生成、鉴赏和演化提供先验基础。

（3）创意生成团队

创意生成团队在艺术知识库的支持下，根据分配的任务流调用生成模型与工具，生成大量创意创作方案^[42]，并通过大规模计算实验对艺术联想、转换和组合等方案进行探索与评估，推荐符合人类

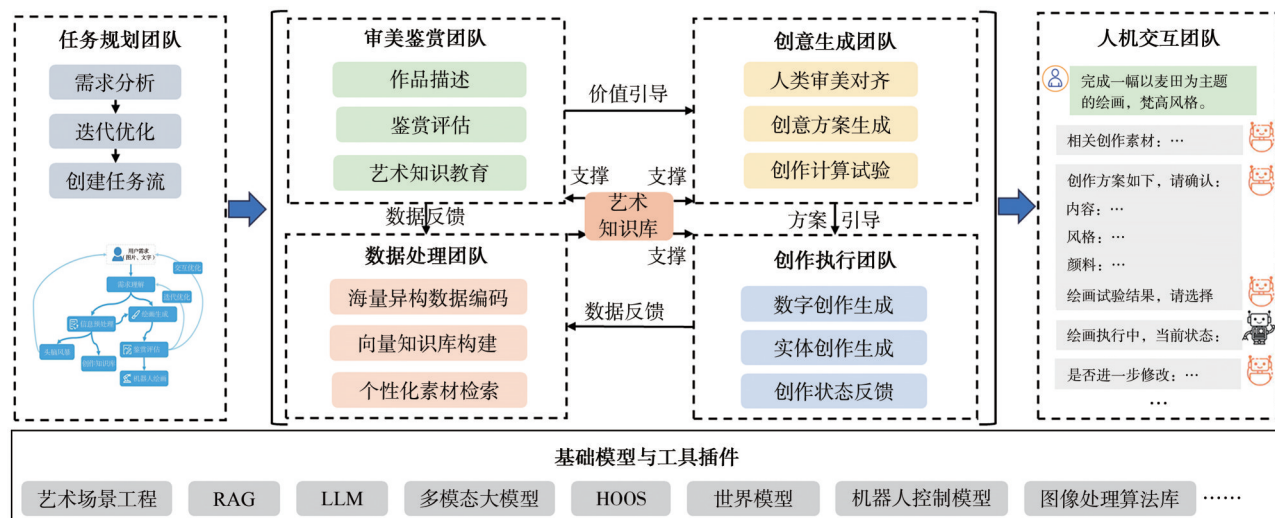


图3 智能艺术工厂的构成

需求和审美要求的创作方案。

(4) 创作执行团队

创作执行团队根据选择的创作方案，调用图像生成模型、机器人控制模型等各类数字和实体的创作工具，在信息空间和物理空间中实现创作过程的自动化，并提供创作状态监控与反馈，进一步通过记忆和反思机制将反馈数据添加至短期记忆数据库和长期艺术知识库。

(5) 审美鉴赏团队

艺术的鉴赏和创作可以认为是互逆的过程，只有具备足够的艺术知识和较强的审美能力，才能够创作出高质量的作品^[43]。审美鉴赏团队主要鉴赏艺术品、学习艺术史，针对艺术创作过程和结果能够给出专业的艺术描述^[44]、创作技能评估和审美价值判断^[45]，一方面为艺术创作智能体提供价值偏好导向，另一方面通过艺术教育提升用户的艺术修养。

(6) 人机交互团队

人机交互团队主要依赖语言大模型、各类多模态采集和展示系统与工具，通过多轮对话和思维链的方式明确人类需求，提供以人为中心的需求理解和引导、创作方案及结果展示、人类示范示教、评估反馈等功能。

(7) 基础模型与工具库

基础模型与工具库根据场景工程对艺术创作过程进行建模和分类，以各类通用多模态大模型和艺术创作领域模型为基础，集成多模态信息处理、世界模型、机器人控制等算法和工具，为各智能体团队提供特定权限的垂直功能。

上述不同角色的数字人团队通过协作配合实现对艺术创作过程的自动分析建模、规划执行、鉴赏评估，按功能动态组合数字人智能体，实现艺术创

作流程自动化，形成以人为中心的艺术工厂操作系统（human-oriented operation system, HOOS），为人类提供艺术创作服务。

3.3 智能艺术工厂的工作流与功能

基于上述智能艺术工厂的底层结构和主要组成，设计艺术工厂的工作流，为用户提供的功能如图4所示。典型的工作流程包括创作需求交互理解、创作方案分析设计、创作任务规划、创作计算实验、创作执行、用户评价、创作结果分析反思、创作知识库积累等关键环节，并基于该基本流程为用户提供艺术创作服务。

智能艺术工厂基于艺术知识库为用户提供艺术知识查询服务。针对特定的用户需求，在基于RAG构建的多模态向量数据库进行检索查询，提供与任务相关的创作知识和数据，为创作任务提供指导；同时，提供与创作需求对齐的作品作为创作方案参考进行头脑风暴；最后，对艺术作品进行分析鉴赏，实现艺术教育。

智能艺术工厂基于计算实验提供人机协同创作服务，根据用户需求和制定的创作方案任务流进行大规模创作计算实验，对不同创作内容、风格、参数进行大量探索实验，并由创作鉴赏团队进行方案评估，将方案及其匹配的评估结果推荐给用户进行评价和确认，最后交付执行。创作实验充分建模艺术创作流程，并根据用户的偏好形成个性化定制的创作知识和创作技能。

3.4 智能艺术工厂的实现方法

为了实现智能艺术工厂及其基本工作流和创作功能，本文采用场景工程进行创作过程建模，并生成用于智能体训练与测试的场景与数据，基于世界模型拟合物理创作规律，从而更准确地建立艺术人工系统，通过RAG增强的多模态大模型学习艺术

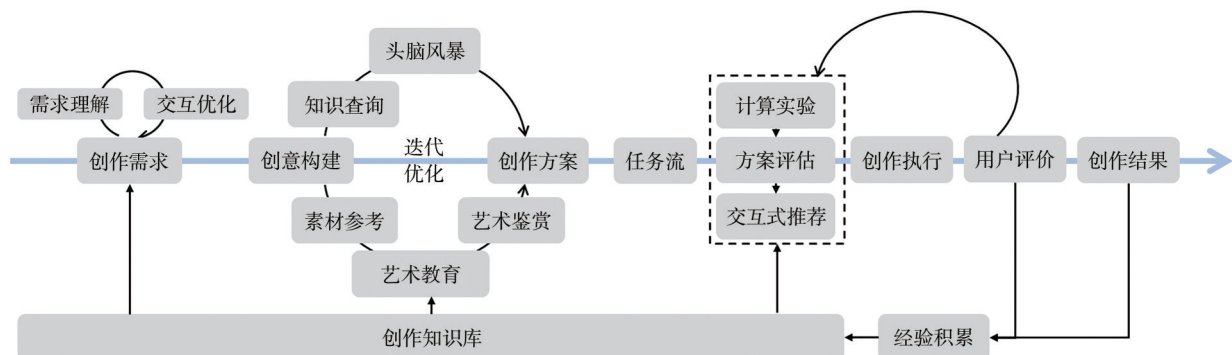


图4 智能艺术工厂的工作流与功能

生成与鉴赏技能,由智能机器人进行实体创作,并通过人类示教、人类反馈等人在环内的方式实现用户个性化定制的创作价值对齐,从而形成以人为中心的艺术操作系统,实现创作流程自动化和创作知识自动化。

(1) 艺术场景工程

艺术场景工程对典型艺术形式、流派、主题、风格和内容进行创作流程建模,基于本体论构建艺术创作的典型场景库。根据不同创作场景,采集和整理人类实际创作过程,建立对应的专家数据库,并进行统计分析和描述学习以建立先验知识库。进一步地,基于场景构建针对功能性智能体的训练数据、测试样例和评估机制,形成针对智能体的艺术训练学校。

(2) 艺术世界模型与仿真方法

人类的创作经验包含了大量对物理创作媒介的认知。对艺术创作物理机理的理解和学习,能够更好地实现人工系统对创作过程的模拟和计算,为拟真的创作计算实验奠定基础。对于传统绘画,建模颜料之间以及与绘画介质之间的交互特性,建立物理特性敏感的环境交互认知模型,能够大幅提高绘画创作计算实验的真实度和推荐创作方案的准确度。绘画过程及其交互特征可以体现在绘画视频中,在一定程度上,视频生成模型为艺术世界模型的建立提供了一种新思路。

(3) RAG增强的艺术大模型

尽管大模型具备强大的语言交互、逻辑推理、多模态融合等能力,但对特定艺术领域知识的掌握不够深入和及时,且无法实现基于隐私数据的创作技能训练与管理。因此,基于RAG^[46]构建艺术大模型,设计多模态预训练任务,学习通用创作与审美技能,并在场景工程的指导下通过模型微调、上下文学习等方式进行专项技能学习。本文在艺术大模型的基础上,面向功能构建智能体和数字人团队,能够满足鉴赏、创作、教育等不同任务需求。

(4) 智能创作机器人

智能机器人是数字艺术家与实际物理环境交互、完成实体创作的重要载体,也是面向人类艺术家的辅助创作工具。然而,要掌握类人的创作技能,需要考虑和建模机器人的物理约束和动作特性。以绘画创作为例,其采用考虑机器人和画笔物理约束的笔触生成方法,规划机器人绘画的笔触序

列,实现了拟人的绘画实体创作^[35,39]。

(5) 人在环内的创作价值对齐

受限于专业精准训练数据的规模和世界模型的拟真度,通过艺术公开数据和数字艺术家在人工系统中的自主探索,仅能学习到广泛的通用创作技能,并仅局限于特定风格,无法面向用户个体提供个性化定制的创作辅助服务。因此,一方面通过人类专家示教数据快速提升数字艺术家和机器人艺术家的专业创作技能。另一方面,通过专家的人类反馈学习典型流派风格,通过用户的人类反馈学习符合用户风格的创作技能,以提供个性化定制艺术服务。

(6) 基于区块链的艺术认证

由于数字人、机器人以人类“分身”的身份在信息和物理空间进行大量艺术学习和创作活动,具有部分或全部的人类艺术创作数据和知识,艺术作品和知识的归属和认证需要通过非同质化代币(non-fungible tokens, NFT)等区块链技术进行明确的鉴定和追溯,并基于智能合约保证合作关系。此外,数字艺术家作为人在信息空间的身份代表,是艺术社区的重要角色,其通过去中心化自治组织(decentralized autonomous organizations, DAOs)组织和管理创意合作活动,实现艺术风格演化^[47]。

(7) 以人为中心的艺术操作系统

智能艺术工厂的最终目标是实现创作流程的自动化和创作知识的自动化,由不同层级和功能的智能体构成的团队为每一个用户提供创作辅助服务,解放大众的创造力。艺术工厂的底层架构、软硬件组成和智能体角色功能,共同构成了以人为中心的艺术操作系统。该系统通过提示工程(prompt engineering)实现对人类真实需求的交互和理解,通过对齐工程(alignment engineering)进行人类价值对齐,通过引导工程(prescriptive engineering)实现人对机器技能的引导以及机器对人创作的建议,最终实现以人为中心的人机协作创作。

4 结束语

现有辅助创作内容生成方法仍然需要专业领域的知识和反复尝试,无法实现创作流程的自动化和以用户为中心的创作知识自动化,难以广泛服务于普通人群。本文基于平行艺术架构和人工智能智能体系统,提出智能艺术工厂的理念和架构,建立与实际艺术创作过程对应的虚拟人工系统,通过生物

人、数字人、机器人跨物理、信息和社会空间的协作设计了艺术工厂的底层架构,构建了艺术创作智能体及其团队,并将其作为艺术工厂的基本构成,阐述了艺术工厂的工作流程、基本功能和实现方法。本文通过智能体员工、数字人团队与人、机器人之间的人机物协同,实现艺术创作流程自动化和创作知识自动演化。继流水线工厂解放生产力之后,智能艺术工厂将进一步解放大众的创造力。

参考文献:

- [1] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 10674-10685.
- [2] LIU Y X, ZHANG K, LI Y, et al. Sora: a review on background, technology, limitations, and opportunities of large vision models[EB]. arXiv preprint, 2024, arXiv: 2402.17177.
- [3] WANG F Y, MIAO Q H, LI X, et al. What does ChatGPT say: the DAO from algorithmic intelligence to linguistic intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 575-579.
- [4] WANG F Y, MIAO Q H, LI L X, et al. When does sora show: the beginning of TAO to imaginative intelligence and scenarios engineering[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(4): 809-815.
- [5] OUYANG L, WU J, XU J, et al. Training language models to follow instructions with human feedback[EB]. arXiv preprint, 2022: arXiv: 2203.02155.
- [6] WEI J, TAY Y, BOMMASANI R, et al. Emergent abilities of large language models[EB]. arXiv preprint, 2022, arXiv: 2206.07682.
- [7] WANG L, MA C, FENG X Y, et al. A survey on large language model based autonomous agents[EB]. arXiv preprint, 2023, arXiv: 2308.11432.
- [8] XI Z H, CHEN W X, GUO X, et al. The rise and potential of large language model based agents: a survey[EB]. arXiv preprint, 2023, arXiv: 2309.07864.
- [9] LI J K, WANG S Y, ZHANG M, et al. Agent hospital: a simulacrum of hospital with evolvable medical agents[EB]. arXiv preprint, 2024, arXiv: 2405.02957.
- [10] GUO C, LU Y, DOU Y, et al. Can ChatGPT boost artistic creation: the need of imaginative intelligence for parallel art[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(4): 835-838.
- [11] WANG B, LI Y L, LV Z Y, et al. LAVE: LLM-powered agent assistance and language augmentation for video editing[C]//Proceedings of the 29th International Conference on Intelligent User Interfaces. New York: ACM, 2024: 699-714.
- [12] TAN W H, DING Z L, ZHANG W T, et al. Towards general computer control: a multimodal agent for red dead redemption II as a case study[EB]. arXiv preprint, 2024, arXiv: 2403.03186.
- [13] QIAN C, LIU W, LIU H Z, et al. ChatDev: communicative agents for software development[EB]. arXiv preprint, 2023, arXiv: 2307.07924.
- [14] HONG S R, ZHUGE M C, CHEN J, et al. MetaGPT: meta programming for A multi-agent collaborative framework[EB]. arXiv preprint, 2023, arXiv: 2308.00352.
- [15] YAO S Y, ZHAO J, YU D, et al. ReAct: synergizing reasoning and acting in language models[EB]. arXiv preprint, 2022, arXiv: 2210.03629.
- [16] WANG L, MA C, FENG X Y, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.
- [17] WEI J, WANG X Z, SCHUURMANS D, et al. Chain of thought prompting elicits reasoning in large language models[EB]. arXiv preprint, 2022, arXiv: 2201.11903.
- [18] LIU R B, YANG R X, JIA C Y, et al. Training socially aligned language models on simulated social interactions[EB]. arXiv preprint, 2023, arXiv: 2305.16960.
- [19] LI G, HAMMOUD H, ITANI H, et al. Camel: communicative agents for “mind” exploration of large language model society[J]. Advances in Neural Information Processing Systems, 2023, 36: 51991-52008.
- [20] LIANG T, HE Z W, JIAO W X, et al. Encouraging divergent thinking in large language models through multi-agent debate[EB]. arXiv preprint, 2023, arXiv: 2305.19118.
- [21] PAUL M, MAGLARAS L, FERRAG M A, et al. Digitization of health-care sector: a study on privacy and security concerns[J]. ICT Express, 2023, 9(4): 571-588.
- [22] PARIKH D, ZITNICK C L. Exploring crowd co-creation scenarios for sketches[EB]. arXiv preprint, 2020, arXiv: 2005.07328.
- [23] DAVIS N, HSIAO C P, YASHRAJ SINGH K, et al. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent[C]//Proceedings of the 21st International Conference on Intelligent User Interfaces. New York: ACM, 2016: 196-207.
- [24] LEE Y J, ZITNICK C L, COHEN M F. ShadowDraw: real-time user guidance for freehand drawing[J]. ACM Transactions on Graphics, 2011, 30(4): 1-10.
- [25] HUANG Z Y, PENG Y C, HIBINO T, et al. dualFace: two-stage drawing guidance for freehand portrait sketching[EB]. arXiv preprint, 2021, arXiv: 2104.12297.
- [26] KARIMI P, REZWANA J, SIDDIQUI S, et al. Creative sketching partner: an analysis of human-AI co-creativity[C]//Proceedings of the 25th International Conference on Intelligent User Interfaces. New York: ACM, 2020: 221-230.
- [27] WANG C, PÉREZ-D’ARPINO C, XU D F, et al. Co-GAIL: learning diverse strategies for human-robot collaboration[EB]. arXiv preprint, 2021, arXiv: 2108.06038.
- [28] LIU H, HU M H, CHEN Y Z, et al. Angel’s girl for blind painters: an efficient painting navigation system validated by multimodal evaluation approach[EB]. arXiv preprint, 2021, arXiv: 2107.12921.
- [29] SUH S, ZHAO J, LAW E. CodeToon: story ideation, auto comic generation, and structure mapping for code-driven storytelling[C]//Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. New York: ACM, 2022: 1-16.
- [30] CHAKRABARTY T, PADMAKUMAR V, BRAHMAN F, et al. Creativity support in the age of large language models: an empirical study involving emerging writers[EB]. arXiv preprint, 2023, arXiv: 2309.12570.
- [31] 王飞跃. 平行系统方法与复杂系统的管理和控制[J]. 控制与决策, 2004, 19(5): 485-489, 514.
WANG F Y. Parallel system methods for management and control of complex systems[J]. Control and Decision, 2004, 19(5): 485-489, 514.
- [32] WANG F Y. Toward a paradigm shift in social computing: the ACP approach[J]. IEEE Intelligent Systems, 2007, 22(5): 65-67.
- [33] 郭超, 鲁越, 林懿伦, 等. 平行艺术: 人机协作的艺术创作[J]. 智能科学与技术学报, 2019, 1(4): 335-341.
GUO C, LU Y, LIN Y L, et al. Parallel art: artistic creation under human-machine collaboration[J]. Chinese Journal of Intelligent Sci-

ence and Technology, 2019, 1(4): 335-341.

- [34] XUE X, YU X N, ZHOU D Y, et al. Computational experiments for complex social systems: integrated design of experiment system[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(5): 1175-1189.
- [35] GUO C, BAI T X, LU Y, et al. Skywork-daVinci: a novel CPSS-based painting support system[C]//Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering(CASE). Piscataway: IEEE Press, 2020: 673-678.
- [36] 倪清桦, 郭超, 王飞跃. 平行戏剧: 新时代戏剧的人机协同创作与智能管理[J]. 智能科学与技术学报, 2023, 5(4): 436-445.
NI Q H, GUO C, WANG F Y. Parallel theaters: human-machine collaborative creation and intelligent management for theatrical art[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(4): 436-445.
- [37] 王春法, 王飞跃, 鲁越, 等. 平行博物馆: 新时代博物馆运营的智能管理与控制[J]. 智能科学与技术学报, 2021, 3(2): 125-136.
WANG C F, WANG F Y, LU Y, et al. Parallel museums: intelligent management and control of museum operations in the new era[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(2): 125-136.
- [38] 鲁越, 郭超, 潘晴, 等. 平行博物馆系统: 框架、平台、方法及应用[J]. 模式识别与人工智能, 2023, 36(7): 575-589.
LU Y, GUO C, PAN Q, et al. Parallel museum systems: framework, platform, methods and applications[J]. Pattern Recognition and Artificial Intelligence, 2023, 36(7): 575-589.
- [39] GUO C, BAI T X, WANG X, et al. ShadowPainter: active learning enabled robotic painting through visual measurement and reproduction of the artistic creation process[J]. Journal of Intelligent & Robotic Systems, 2022, 105(3): 61.
- [40] GUO C, DOU Y, BAI T X, et al. ArtVerse: a paradigm for parallel human-machine collaborative painting creation in metaverses[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2200-2208.
- [41] LU Y, GUO C, DAI X Y, et al. ArtCap: a dataset for image captioning of fine art paintings[J]. IEEE Transactions on Computational Social Systems, 2024, 11(1): 576-587.
- [42] LU Y, GUO C, DOU Y, et al. Could ChatGPT imagine: content control for artistic painting generation via large language models[J]. Journal of Intelligent & Robotic Systems, 2023, 109(2): 39.
- [43] LU Y, GUO C, LIN Y L, et al. Computational aesthetics of fine art paintings: the state of the art and outlook[J]. Acta Automatica Sinica, 2020, 46(11): 2239-2259.
- [44] LU Y, GUO C, DAI X Y, et al. Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training[J]. Neurocomputing, 2022, 490: 163-180.
- [45] LU Y, GUO C, DAI X Y, et al. Generating emotion descriptions for fine art paintings via multiple painting representations[J]. IEEE Intelligent Systems, 2023, 38(3): 31-40.
- [46] DAI X Y, GUO C, TANG Y, et al. VistaRAG: toward safe and trustworthy autonomous driving through retrieval-augmented generation[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(4): 4579-4582.
- [47] WANG F Y. The DAO to MetaControl for MetaSystems in metaverses: the system of parallel control systems for knowledge automation and control intelligence in CPSS[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(11): 1899-1908.

[作者简介]



郭超 (1992-), 男, 博士, 中国科学院自动化研究所助理研究员, 主要研究方向为机器学习、深度学习、机器艺术创作、智能机器人系统。



戴星原 (1993-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统国家重点实验室助理研究员, 主要研究方向为平行艺术、平行学习、平行智能。



廖玮琰 (2003-), 男, 北京科技大学本科生, 主要研究方向为AI智能体、智能机器人系统。



祁珏瑜 (2002-), 男, 北京科技大学本科生, 主要研究方向为图像生成模型、扩散模型。



鲁越 (1994-), 男, 博士, 山东大学控制科学与工程学院博士后, 主要研究方向为机器学习、小样本学习、机器人系统。



倪清桦 (1999-), 女, 澳门科技大学博士生, 主要研究方向为平行智能、数字孪生、智能系统。



Wang Lin (1976-), 女, 博士, 德国柏林德意志歌剧院、奥地利格拉茨歌剧院歌剧作曲家, 加州大学伯克利劳伦斯国家实验室交叉学科研究员, 南京艺术学院客座教授。主要创作方向为歌剧及交响乐作曲、剧作家、戏剧构作, 主要研究方向为艺术与科学交叉、超学科研究和创作。



王飞跃 (1961-), 男, 博士, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任, 兼任中国国家博物馆科学与艺术研究所所长。主要研究方向为平行系统的方法与应用、社会计算、平行智能、知识自动化。