大模型时代下生成式可视媒体的机遇与挑战
——第二十三期 CCF 秀湖会议报告

整理：杨 鑫[1] 王贝贝[2] 过 洁[2] 夏佳志[3] 王莉莉[4] 吕 琳[5] 刘利斌[6]
陈雪锦[7] 高 林[8] 赫 然[9] 徐 凯[10] 周 昆[11]
[1] 大连理工大学 [2] 南京大学 [3] 中南大学 [4] 北京航空航天大学 [5] 山东大学
[6] 北京大学 [7] 中国科学技术大学 [8] 中国科学院计算技术研究所
[9] 中国科学院自动化研究所 [10] 国防科技大学 [11] 浙江大学

背景

可视媒体涵盖了图形、图像、视频以及虚拟现实（virtual reality, VR）、增强现实（augmented reality, AR）等各种基于视觉的内容，是人类视觉认知的核心载体，也是数字世界中最关键的信息表达形式。党的二十大报告提出，要“加快发展数字经济，促进数字经济和实体经济深度融合，打造具有国际竞争力的数字产业集群”。可视媒体作为数字经济的重要组成部分，在智能制造、社会民生、数字文创、智慧城市、国防安全等领域广泛应用，为提升经济效率和优化资源配置提供了新动力。

可视媒体技术通过采集、编码、处理、生成与交互等过程，实现人–机–物世界的数字化表达与理解。生成式人工智能推动了可视媒体生成新范式的形成，以 Sora、Stable Diffusion、Vidu、可灵 AI、通义万相等图像视频生成技术为代表，“语言–视觉”双模态的生成方式极大地提升了内容产生、交互表达和空间理解能力，颠覆了传统技术，引发了各国政府和行业巨头的高度关注和积极布局。

然而，当前的生成式可视媒体技术仍面临关键技术瓶颈，主要包括：缺乏物理规律约束、空间几何一致性差、交互控制精度不足、可信鉴权与溯源机制薄弱、生成性能与能效比低、触/味/嗅扩展媒体形态生成研究空白等问题。这些问题严重制约了生成式可视媒体技术在智能制造、自动驾驶、虚拟仿真、沉浸交互等领域的应用。因此，亟须开展面向三维物理世界的生成式可视媒体的计算理论与方法的基础研究，推动“数据驱动”与“物理驱动”的深度融合，实现从内容生成走向物理认知与推理智能的范式跃迁，支撑国家在新一轮人工智能技术变革中的战略竞争力。

在此背景下，2024 年 11 月 22—24 日，第 23 期中国计算机学会（China Computer Federation, CCF）秀湖会议在苏州 CCF 业务总部&学术交流中心召开。会议设置五大专题：**生成式可视媒体理论与计算架构**，探讨混合表达范式、国产计算框架（如计图）与专用芯片的协同创新；**生成式图形与计算机辅助设计（computer-aided design, CAD）**，聚焦统一几何表达、三维高斯泼溅、物理约束嵌入、工业生成模型与柔性制造工艺优化；**生成式虚拟现实**，探索多感官一致性融合、增强现实交互情景化、轻量化生成与移动端生态构建；**生成式影像**，高保真可控生成、图形–图像跨模态协同、时空一致性约束等关键技术；**Beyond 生成式可视媒体**，在脑科学、影视创作等跨学科领域的应用，并探讨应对伦理与

DOI：10.11991/cccf.202510009

内容安全的挑战。

会议系统梳理了“生成式可视媒体”的技术瓶颈并达成行动共识：一方面，确立生成式可视媒体计算框架、三维原生特征建模、物理–几何耦合表达、多模态认知语义对齐为理论攻坚重点；另一方面，联合发起“构建开放共享的工业级 CAD 等三维数据集、建立跨学科可视媒体生成质量评估标准”等倡议，推动产学研协同创新。会议在大模型驱动的可控性可视媒体计算新范式、多感官 VR 内容生成技术新体系、高可控性图形–图像协同生成新机制、内容生成技术的伦理与安全治理新框架等方向达成共识，总结了“**生成式可视媒体”的技术瓶颈、核心挑战与发展路径**，为生成式可视媒体的创新技术突破与重要产业落地锚定方向。

生成式可视媒体的物理–语义统一表示机制 针对当前生成模型侧重语义信息建构，缺乏对物理属性和空间结构的准确表征，难以实现真实可信的生成与高精度交互控制的核心科学问题，需研究构建融合视觉原生特征、空间几何拓扑、物理状态约束的统一可视媒体表示模型。探索多模态大模型中的语义–几何–物理对齐机制。发展面向物理过程的可微表示学习方法，提升模型对真实世界复杂场景的建模与理解能力。

多模态可视内容的高可控生成与沉浸式表达 针对当前生成式系统控制粒度粗、响应延迟高、缺乏场景一致性，难以满足沉浸式交互、协同感知等实际应用需求的核心科学问题，需研究高可控、多粒度、多感官协同的生成机制，支持“文本—图像—视频—声音—触觉”等多模态数据的动态生成。实现情境感知增强现实系统的自适应演化机制。攻克轻量化、低延迟、高保真 VR/AR 内容生成与传输关键技术，支撑实时沉浸交互场景。

面向物理逻辑与时空认知的可视媒体智能推理机制 针对当前多模态生成系统虽能合成丰富的视觉内容，但缺乏对场景中物理因果关系、时空演化规律和行为逻辑的理解与推理能力，难以支持具身智能、智能交互与预测性决策等高阶任务的核心科学问题，需构建融合时空结构建模与物理规则嵌入的认知表示体系，实现三维动态场景的语义理解与行为推理；发展基于可微物理表示的大模型驱动的动态推演与预测机制，支持真实环境中连续、可控的事件生成。研究因果推理与反事实生成技术，提升系统对复杂场景中行为后果的解释与预判能力。实现认知与生成一体化架构，推动生成式可视媒体系统从内容表达走向空间理解与自主响应，支撑智能体、虚拟人和复杂交互系统的智能行为建构。

沉浸式可视媒体的交互感知–认知计算 探究包括可视媒体视、听、触等多模态信息感知的融合机制及其交互自然性评价指标体系，需构建沉浸式人机交互的底层感知–认知模型。提出可计算心理模型，准确刻画人类与可视媒体及智能终端在不同交互情境下的决策机制。通过揭示人类多情境认知机理，模拟人类情境化决策过程，为建立思维层面人机协同系统提供理论支持与技术基础。开发支持多模态信息呈现与空间计算的终端设备，构建可感知和理解现实环境的虚实融合交互空间，最终在虚拟世界中实现与现实世界无缝衔接的互动体验，显著提升用户沉浸感。

可视媒体生成的安全治理与工业应用 针对当前生成式系统普遍缺乏可信生成机制，缺乏物理一致性与工程适应能力，制约其在制造、自动化等场景中的可用性的核心科学问题，需构建具备物理一致性约束与可追溯鉴权能力的生成式内容安全治理机制。实现面向制造业的可视媒体驱动产线孪生、工艺流程仿真与虚实同步演化技术。研究嵌入物理建模约束的复杂制造流程快速适配、精细优化与闭环决策机制，支撑智能制造系统升级。

生成式可视媒体计算框架 针对当前生成式架构多为数据驱动，难以表达连续时空物理规律，且计算开销大，难以适配工业级场景与空间智能系统需求的核心科学问题，需构建融合物理几何规律与可微建模能力的生成式神经计算新框架。研发适用于 CAD/计算机辅助工程（computer-aided engineering, CAE）的空间智能可视媒体计算引擎，实现可控、可复用、可解释的智能设计协同。提出支持跨平台异构部署与虚实闭环反馈优化的大模型计算体系，推动从“内容生成”向“空间理解与决策”的范式转变。

五大专题论坛

生成式可视媒体理论与计算架构

生成式 AI 技术通过双模态生成提升创作效率，推动行业商业化（如可灵 AI 单月流水超千万元）。专家

共识包括：

1）**混合表达范式** 需融合传统几何简洁性与 AI 计算高效性，更准确地表达边界、拓扑、物理等属性。

2）**特征嵌入空间优化** 结合图形知识提升特征嵌入的可理解性和操纵性，探索二维特征控制三维生成。

3）**高效计算架构** 发展国产框架（如计图）与专用芯片（如 Nerf 渲染芯片），适配工业级场景需求。

4）**大语言模型融合** 利用大语言模型的知识表达能力，推动人机交互和可视分析向“感人、知人、拟人”的方向发展。

生成式图形与 CAD

生成式 AI 赋能图形与 CAD，提升设计制造效率与质量，推动系统集成协同及全流程智能化以满足智能制造需求，专家共识包括：

1）**统一数据表示与计算框架** 构建支持参数化、隐式、体素表达无缝切换的混合框架，兼顾精确性与高效性，实现跨平台协同设计。

2）**物理约束嵌入** 在生成式 AI 中引入力学、热学等多物理场约束，形成物理可控的 CAD/CAE 生成范式，满足高精度仿真需求。

3）**大规模 CAD/CAE 数据集构建** 解决现有数据集规模小、精度低问题，建立开放共享的高质量数据集，支撑模型训练与评估。

4）**复杂制造工艺优化** 融合生成式模型与柔性制造，实现工艺自动设计、仿真优化，如 3D 人工智能生成内容（artificial intelligence generated content, AIGC）、产线孪生、工艺流程仿真。

生成式虚拟现实

生成式人工智能应用于虚拟现实，重新定义技术边界并开辟高效、高质量、可控体验新道路，VR/AR 技术需突破沉浸感与交互性瓶颈，专家共识包括：

1）**多感官一致性融合** 融合视觉、听觉、触觉等多通道传感器信息，研发高真实感 VR 内容生成方法，降低生产成本。

2）**自然交互生成** 优化对象—人—智能体交互模型，引入物理仿真提升行为真实性，利用机器学习增强交互自然性。

3）**轻量化与实时化技术** 优化生成网络与渲染算法，减少计算资源依赖，适配移动端设备，推动 VR 普及。

4）**计算平台与终端研发** 构建支持大规模数据处理与实时生成的计算平台，开发高效终端设备，完善 VR 生态。

生成式影像

生成式影像技术的发展，为理解人类认识世界的方式提供了新的视角，高清晰、高保真、时空一致的影像是产业落地核心需求，专家共识包括：

1）**高质量影像表征** 通过多模态交互，充分提升并利用大模型在物理理解和推理方面的能力以应对幻觉问题。

2）**高可控生成** 结合图形学三维建模优势，实现“文本—图像—视频”多模态控制，在表达、建模、可控性等方面实现优势互补。

3）**图形–影像跨模态协同** 研究时空约束下的影像生成（如三维纹理生成），满足 CAD/CAE、VR/AR 等领域需求。

Beyond 生成式可视媒体

生成式可视媒体技术能以较低成本高效产生带标签和关联的高质量图形图像，推动跨学科研究范式变革，围绕“Graphics+X”方向，专家形成的共识包括：

1）**跨领域数据生成** 把图形看作仿真引擎，生成式可视媒体技术可应用在非可见光感知、具身智能、工业生产、智能制造、元宇宙等领域。

2）**专业知识融合** 结合特定领域知识，设计科学的生成质量评估指标，构建规范的真实性、可靠性度量和控制。

3）**跨学科协作** 基于图形学理论基础，探索生成式可视媒体技术在其他领域应用中的本质科学问题，不断扩宽边界，扩大影响。

生成式可视媒体的十二大科学技术问题

围绕 5 个专题内容，与会专家开展了深入讨论，凝练出生成式可视媒体的 **12** 个科学技术问题，如图 1 所示。

图1　生成式可视媒体的 12 个科学技术问题

生成式可视媒体的表示机制　在图形、图像、视频等可视媒体生成任务中，高效地描述可视媒体中对象的几何、材质等信息，为生成、渲染、编辑和交互提供数据表达基础。可视媒体表示涉及 3 个关键问题。第一，任务普适的可视媒体表示。这种表示对生成、渲染、编辑和交互的计算机制和计算效率起到决定性的作用。反过来，生成、渲染、编辑和交互等任务对可视媒体表示的要求具有差异性。高效训练和推理、适配工业图形流水线、易于理解和编辑等要求对表示的普适性提出了挑战。构建任务普适、紧凑高效的可视媒体表示，可为生成式可视媒体提供基础条件。第二，三维可视媒体与源数据维度不对等。一方面，三维图形数据的采集能力远弱于二维图像数据，导致现有三维图形数据的规模难以支持通用生成任务；另一方面，在可视媒体生成任务中，实现几何语义约束需要三维表示。高效获取三维数据并得到原生三维表示，或者从二维数据中高效提取三维语义信息，是实现高质量三维可视媒体生成的关键。第三，融合几何与物理属性的统一表示。如何在生成式模型环境下，构建几何信息与力学结构性能等高阶属性的统一表示，是重要的研究方向。总之，表示机制是生成式可视媒体的关键问题，决定了生成式模型的计算机制和适用范围。

生成式可视媒体的内容控制　生成内容控制通过介入生成过程，交互定制可视媒体生成内容，以精准满足任务需求。生成式可视媒体的内容控制涉及 3 个关键问题。第一，多模态控制信息与可视媒体表示对齐。文本、图像、图形、草图等不同模态的控制输入包含了丰富而独特的语义信息，起到增强可视媒体内容描述的作用。尤其是来自大语言模型的文本控制信息，具有较大的内容描述能力。如何将多模态控制信息与可视媒体表示对齐，是生成内容控制的关键问题。第二，可调控的特征空间。当前的生成式模型多采用编码器–解码器架构。在这个架构下，生成内容调控在特征空间中发生作用。如何提高特征的表达能力和可调控能力，是可控生成的重要问题。第三，可控生成的质量评估。可计算的质量评估指标是高质量可控

生成的基础。控制信息的达成度评估、生成内容的质量评估，以及通用的评估数据集建立，都是重要的研究内容。总而言之，通过研究生成式可视媒体的内容控制，可以为可视媒体生成提供可交互、可定制、高质量的解决方案。

生成式可视媒体的计算架构 可视媒体生成在历史上沿着二维和三维两条不同的路线发展。二维的计算架构主要基于图像、梯度和频域的操作，以满足二维表观需求。三维的计算架构主要基于渲染框架，遵循图形学特有的物理规律简化原则，以满足真实高效的需求。随着人工智能的快速发展，图形学和视觉的边界逐渐模糊，呈现二维和三维融合、知识体系融合的重要趋势。二维的计算架构，如果其内在表达和分析完全脱离三维，将难以满足三维中的一致性和合理性要求。三维的计算架构，其难点在于生成渲染流水线需要的高保真几何、材质、光照等输入信息。因此，未来生成式可视媒体的一个关键科学问题是融合二维和三维路线优势的计算架构。这需要考虑 3 个方面的问题。第一，计算架构的内部表示至关重要。它既需要能利用丰富的二维数据，也需要能全面地描述三维信息。简化的三维表示或混合表示是值得探索的方向。第二，计算架构需要跳出传统遵循物理规律简化的渲染流水线，又需要符合物理规律。神经渲染架构如何融合物理规律，是值得思考的问题。第三，可视媒体生成的可编辑性和可控性对计算架构的可微性提出了内在要求。

融合几何表达与 AI 计算的统一框架 随着 CAD/CAE 系统对高效性、精确性及可扩展性的需求提升，构建生成式可视媒体的统一几何表达与计算框架成为核心挑战。该框架需高效处理复杂几何结构，包括高拓扑复杂性、多分辨率表达及动态几何变化，以降低计算资源消耗并提升生成效率。传统几何表达中，参数化方式适用于高精度设计，隐式表达适于复杂形状生成，体素化表达利于数值计算。为整合各自优势，框架须支持混合表达间的无缝切换，使用户可根据设计与计算需求灵活选择。此外，框架须具备跨平台协同能力，实现多 CAD/CAE 系统间数据与计算的集成共享，支持多用户协同设计与优化。同时，还要解决异构数据兼容性问题，提升灵活性与适配性，以满足智能设计与制造需求。其中的关键问题包括以下 5 个方面：第一，统一框架需兼顾表达简洁性与计算高效性。传统几何表达精确但计算复杂，AI 方法高效却缺乏精度，需融合两者优势。第二，多尺度复杂几何建模是关键。框架需处理高拓扑与动态变化，支持从宏观设计到微观优化的多分辨率表达。第三，混合表达切换需无缝高效。参数化、隐式与体素表达各有局限，框架需实现无损转换与实时适配。第四，跨平台协同需数据与计算一体化。异构系统间的数据流转与计算共享需统一标准支持多用户协作。第五，异构数据兼容性需系统性解决。框架需整合不同来源的数据格式，提升适配新一代智能制造的灵活性。

在生成机制中嵌入物理约束 生成式 AI 在 CAD/CAE 及智能制造中的应用日益深入，尤其在航空航天、能源装备等领域，但现有模型在物理功能的生成与可控性上发展受限。核心挑战在于将生成机制与多物理仿真融合，通过嵌入物理约束提升精度与可控性。生成模型需输出满足几何有效性、多物理性能及制造可行性的结果，包括复杂外形以及内部结构。模型需集成力学、热学等多场耦合约束，适应高端制造场景，并利用实时仿真反馈优化设计，推动智能制造发展。其关键问题包括以下 5 个方面：第一，生成需统一几何与物理一致性。现有模型偏重表观几何，物理功能可控性不足，需关联模型确保功能性。第二，多物理场耦合需全面建模。传热与刚度交互复杂，如内部构型影响性能，需细粒度映射提升可控性。第三，生成过程需多约束协同优化。现有模型难以精确控制物理性能，需构建物理参数与生成模型的可微映射，利用梯度驱动优化实现目标导向生成。第四，实时反馈需融入的制造约束。仿真与加工需要动态调整输出，确保物理功能的可行性。第五，复杂场景适配需突破局限。模型需在多场、多尺度约束下增强物理功能的精度与可控性。

复杂制造工艺的快速适配与实时优化 制造业正转向“多品类、小批量”的柔性化生产模式，相较于传统大批量标准化生产，该模式要求制造系统快速响应多样化需求，实现工件间的工艺切换与优化。然而，复杂

制造工艺涉及多物理场耦合、高维参数调控及动态环境扰动，传统依赖人工经验或手动调整的方法难以满足高效性与精确性要求。基于大模型的智能制造系统应实现工艺参数的自动化生成、动态校准与实时优化，推动柔性化生产的智能化发展。该“工艺大模型”利用工件几何特征、材料属性及生产目标生成加工路径、速度分布和温度场控制策略，通过实时多模态数据反馈优化工艺过程，提升生产质量与资源利用率。其核心问题包括以下 5 个方面：第一，柔性化生产依赖工艺快速适配能力。传统方法通过经验或迭代试验确定参数，难以适应多品类工件切换。工艺大模型基于预训练知识与实时输入，迅速生成初始工艺参数。第二，参数自动化生成需融合多源数据与多物理场仿真。复杂工艺涉及机械、热学及流体力学耦合，单一模型难以精确描述。工艺大模型集成仿真分析与生产数据，构建高维参数映射，实现特性解析与结果预测。第三，实时优化需解决动态扰动问题。制造中环境变量与工件状态持续变化，初始参数易偏离最优。模型通过多模态实时数据驱动闭环调控，确保工艺稳定性与最优性。第四，多品类适配要求模型通用性与特异性融合。柔性生产的多样性对传统专用工艺带来挑战。工艺大模型通过模块化设计与少样本学习，支持任务间快速切换，降低适配成本。第五，模型需适配分布式制造与人机协同特性。现代生产涉及多设备并行与多工序协同，模型整合分布式数据流实现全局优化，并通过交互接口支持工程师决策。

全感官一致与耦合的生成　虚拟现实技术的终极目标是实现全感官的沉浸式体验，虽然视觉与听觉的呈现已取得显著进步，但触觉、味觉、嗅觉等其他感官的融入仍面临严峻挑战。当前，不同感官内容往往独立生成，缺乏必要的同步与协调，这种不一致性严重削弱了用户在虚拟环境中的沉浸感受与真实体验。为了进一步提高沉浸感，需要解决的一个关键问题是如何构建高一致性、深度耦合的全感官内容表征。构建这样的表征需要考虑 4 个问题。第一，多种感官信息的同步与校准。不同感官内容的生成需要在时间上精确同步，以确保用户感知到的是一个连贯且协调的虚拟世界，这要求系统能够实时调整和处理来自不同感官源的信息，消除任何可能的延迟或错位，从而实现视觉、听觉、触觉、味觉和嗅觉等感官内容之间的无缝衔接。第二，感官交互的自然性与流畅性。全感官虚拟现实体验的核心在于用户与虚拟环境的自然交互，因此，必须设计出能够模拟真实世界中感官交互特性的系统，包括触感的真实性、嗅觉的细腻度以及味觉的多样性等。此外，这些交互方式应尽可能减少用户的认知负担，使操作直观易懂，从而提升用户体验的自然性和流畅性。第三，多模态感官数据的融合与处理。为了实现高一致性的全感官内容表征，需要将来自不同感官的数据进行有效融合和处理，这要求开发能够处理复杂多模态数据的算法和技术，能够准确识别、解析和整合来自不同感官源的信息，从而生成一个统一且一致的虚拟感官体验。第四，硬件与软件的协同优化。构建高一致性、深度耦合的全感官内容表征不仅需要先进的软件算法，还需要高性能的硬件支持，因此，必须注重硬件与软件的协同优化，确保硬件能够准确、高效地捕捉和呈现多模态感官信息。

物理规律约束的虚拟环境可控生成　虚拟环境的物理真实性对于虚拟现实至关重要。它不仅影响着用户的沉浸感和体验质量，还是衡量虚拟现实技术成熟度的重要指标。一个具有高度物理真实性的虚拟环境，能够让用户感受到更加逼真的场景和交互，从而提升其参与度和信任感。因此，在虚拟现实技术的发展过程中，需要解决的一个关键科学问题是可控生成具有物理真实性的虚拟环境，实现高品质虚拟现实体验。为了保证虚拟环境真实性，需要考虑 3 个问题。第一，生成式人工智能正在快速革新虚拟环境的生成方式，通过深度学习技术可以合成具有高真实感的虚拟数字内容。然而，当前的虚拟环境生成方法存在可控性差的问题，难以精确定义几何、材质和物理属性，导致生成的虚拟环境缺乏三维一致性和物理真实性。通过文本或其他模态能够更精确定义生成的虚拟环境的几何形状、材质属性和动态物理属性。同时，可以加强引入三维生成模型与几何规则先验，以增强生成虚拟环境的三维一致性。此外，还需要在模型中显式定义物理参数并进行仿真，以提升生成虚拟环境的物理真实性。第二，虚拟现实中的虚拟角色和虚拟代理与

用户的交互行为往往缺乏自然性和逼真感，存在表情、动作和语言机械化，情感反馈失准等问题，影响沉浸感和互动体验。为了提升逼真度，需要研究深度学习驱动的行为生成技术，建立情感状态到行为动作的映射模型，实现情感的可控调节，使虚拟角色的行为更加符合人类的行为习惯和情感反应。同时，加强精确的面部表情和肢体动作捕捉技术的研究，并通过生物力学仿真确保动作轨迹符合物理规律，以使虚拟角色的表情和动作更加自然流畅且真实。此外，还需要进行多模态情感与意图推理的研究，以更好地理解用户的意图和情感，从而提升虚拟角色的交互体验。第三，虚拟现实中的动态环境生成与更新需要大量的计算资源，尤其是在生成复杂的三维虚拟环境和动态交互元素时，实时计算需求超出平台处理能力，导致动态场景更新缓慢。为了提升计算效率，可以考虑结合边缘计算和云计算技术，实现计算资源的动态分配和优化。同时，加强高效的图形渲染与分布式渲染技术的研究，以提升渲染速度和质量。此外，还需要对物理引擎和内容生成算法进行优化，以降低计算复杂度，提升生成效率。

高质量影像表征与影像生成　高清晰度、高保真度、时空一致且逻辑合理的影像生成是产业落地与技术突破的核心需求。这一问题面临三大关键问题。第一，如何构建精准、全面且隐含三维特征的影像表征，使模型能够理解复杂的视觉信息，捕捉场景细节，并融入三维时空约束，是生成高质量影像的核心难题。影像表征的学习不仅需要关注局部像素特征，还需提升模型对全局空间关系的理解能力，使影像生成符合人类视觉认知和物理规律，从而确保生成结果在各个维度的协调一致。第二，生成影像的逻辑合理性与时空一致性问题。影像生成过程往往面临长时序动态的不稳定性，生成模型在时间维度上的连贯性不足，容易导致不合理或不自然的影像过渡，出现物体扭曲、遮挡关系错误或动态影像中物体跳动等问题。如何在影像生成过程中引入时空约束和因果约束，确保生成影像在长时间序列中保持稳定性和逻辑性，是实现高质量影像的关键问题。第三，生成结果的评估问题。评估影像生成结果的质量与合理性对于模型优化和实际应用至关重要。如何建立有效的评估机制，量化影像生成的真实感、语义一致性与逻辑合理性，使模型能够自动识别生成影像中必要的细节特征，同时过滤冗余或错误特征，对生成结果进行自适应优化与动态修正，确保影像生成结果符合预期标准并具备较高的真实感，是高质量影像生成的重要问题。

影像生成的高可控性与精细化生成　如何在生成影像的同时实现细粒度可控性，以满足专业领域对复杂细节的精确调整和动态需求，是生成式影像技术发展的核心挑战之一。高可控影像生成涉及 3 个关键问题。第一，细粒度控制与可解释性问题。生成模型在复杂场景下难以精确控制局部细节和微观特征，导致生成结果在空间结构、光影变化以及动态细节方面存在偏差。如何在影像生成过程中引入多尺度、多层级的控制机制，使用户能够在影像生成过程中实时调整和精细雕刻影像的局部特征，确保生成结果在整体布局和细节层面均符合预期，是提升影像生成质量的关键。第二，专家知识与物理约束的融合问题。当前数据驱动的影像生成方法在细节复杂、物理约束强的场景下存在泛化性不足的问题，容易导致生成影像缺乏真实感或在动态变化中失去一致性。如何将三维建模、动作分析和动态物理仿真等领域的专家知识嵌入影像生成模型，通过引入精确的空间约束和物理规律，确保生成影像在细粒度控制下具备物理真实感和合理性，是细粒度影像生成的核心挑战。第三，多模态交互与渐进式细节完善问题。影像生成过程需要结合图像、文本、语音等多模态输入，实现逐步调整和细粒度控制，使用户能够对生成影像的局部细节和整体风格进行交互式优化。如何设计灵活高效的多模态交互机制，使用户能够动态微调影像特征、增强局部细节或调整空间结构，确保影像生成结果在不同尺度上保持一致性和高保真度，是提升影像生成可控性的核心问题。

生成式可视化中的人机协同　生成式可视化的显著特征是生成式模型贯穿可视化的生产与使用过程，在全流程中协同人与机器的计算、感知、认知、规划和决策能力，实现可视化工具在线开发与使用流程的融合。这需要考虑 3 个问题。第一，用户表达与机器表示之间的对齐策略。人对任务需求、设计需求和探索决策的描述往往处于较高的语义层次，而机器的表示则需要处于可以被精确执行的较低语义层次，两者之

间存在知识、经验、探索上下文等语义鸿沟。人的描述往往还包括歧义和不确定性。如何将用户表达与机器表示对齐，是生成式可视化中人机协同的关键。第二，人机任务分配策略。在探索性可视分析过程中，对任务和数据的认知是逐步形成、迭代上升的。人的判断力和创造力起到主导的作用。同时，大模型的计算能力、知识能力、规划和决策能力可以起到重要补充甚至是局部主导作用。如何合理分配任务，是重要的研究方向。第三，基于生成式模型的可视推理。传统交互可视推理范式结合了可视表达的数据洞察能力和机器的计算能力，但分析效率仍然受限于用户。如何构建以可视化为基础表达的生成式可视推理模型，实现自动化的类人可视推理，是值得研究的问题。总之，研究贯穿可视化生产与使用全过程的人机协同机制，有助于高效完成复杂可视分析任务，催生可视化工具生产与使用的新范式。

多模态、多尺度、非结构化数据生成质量评估和反馈　无论是影视制作、数字孪生等传统领域的可视媒体生成，还是 AlphaFold 所代表的在科学领域内三维结构等可视数据的生成，AIGC 技术都在不断地推动工业生产和科研范式变革。然而，不同应用领域都有各自的数据特点和系统逻辑，这些不同模态、不同尺度数据呈现出巨大差异，且数据之间深度耦合，关系复杂，具有高度非线性的特点。而在智能化科研（artificial intelligence for Science, AI4Science）等领域，对内容生成的要求也从高逼真度扩展到了可分析、可解释，以辅助科学探索和发现。因此，需要探索多模态、多尺度、非结构化复杂数据生成质量的科学评估方法和高效反馈机制，进行系统化战略规划，建立数据和算法共享平台和体系，以构建坚实可靠的数据底座和分析平台。

生成式可视媒体的五类应用场景

结合生成式可视媒体的发展前沿和符合中国国情的原则，与会代表讨论并总结了生成式可视媒体的五类应用场景，如图 2 所示。

图 2　生成式可视媒体的五类应用场景

工业制造 制造业是国民经济的主体，是立国之本，智能化、柔性化制造是新型工业化的重要发展方向。柔性智能制造面临诸多挑战，包括柔性产线构建开发成本高、制造需求变化工艺更新慢、产线排产调度优化难、复杂制造工艺在线控制难等。生成式 AI 在海量知识整合、多模信息融合、物理世界建模、长程任务规划、交互动作合成等方面的优势，为智能柔性制造的发展带来了新动力。例如，基于生成式 AI 的产品设计与优化，基于生成式 AI 的制造工艺自动生成、仿真优化，基于多模态大模型的复杂工艺自主决策、最优控制等。

建筑设计 建筑设计是一项复杂耗时的工作，建筑设计中的三维模型是个必不可少的交付件。而对于一些复杂几何形状和异形结构，三维建筑模型的重建工作难度极大。生成式 AI 与建筑信息模型（building information modeling, BIM）软件结合，可以实现只需要一些样本和文字，便可自动创建各种风格的建筑模型，提高建模效率。同时，还可以赋予建筑设计不同的材质，来完善设计。此外，生成式 AI 也可以管理和分析 BIM 中的大量数据，支持决策制定。

教育培训 生成式 AI 可以通过创造性地生成图像、视频、动画等视觉内容，为教育培训行业提供多种服务。我们可以根据学生的学习风格和进度，生成个性化的教材、图表和演示文稿。也可以创建定制化的教学视频，其中虚拟教师可以针对学生的特定需求进行讲解。此外，生成式 AI 可以生成 VR、AR 教学场景，提供沉浸式学习体验，模拟实验、历史场景或者复杂概念，可以让学生在安全的环境中进行实践和学习。

影视娱乐 影视娱乐行业是目前生成式 AI 潜在应用最多的行业。在角色设计方面，生成式 AI 可以设计角色的外观，包括面部特征、服装和道具，甚至可以根据角色的性格特点自动生成相应的形象，依据故事情节和导演意图演绎人物的性格形成与成长过程。也可以更为生动地展示动态人物场景，例如数字人可以逼真展示人物的音容笑貌、行为姿态。在场景生成方面，生成式 AI 可以生成气势恢宏的动画场景，例如栩栩如生的动态自然景观、惨烈战场景象，节省搭建实体场景的时间和成本。在特效制作方面，生成式 AI 可以用于制作电影中的特效，如爆炸、流体模拟、人群模拟等，提高特效的真实感和制作效率。

医疗健康 利用生成式 AI，我们可以显著提升医学图像的质量，并从二维切片图像中重建出整个器官的三维模型。基于这些模型，可以为患者定制个性化的三维解剖模型，从而协助医生进行精准的手术规划和模拟。在手术前，医生还可以借助生成式 AI 进行虚拟手术，以评估和选择最佳手术方案。此外，结合 VR、AR 技术，我们能够使患者直观地了解手术过程或疾病状态。在康复阶段，生成式 AI 还能创建虚拟环境，通过 VR 和 AR 技术辅助患者进行康复训练。

生成式可视媒体技术对我国技术突破与产业落地的推动作用

近年来，生成式人工智能的兴起，尤其是以扩散模型、生成对抗网络（generative adversarial network, GAN）与多模态 Transformer 为核心的可视内容生成能力，正引领全球人工智能和视觉计算领域的范式转变。我国在“以科技自立自强引领高质量发展”的战略指导下，急需一批具备原创性、通用性与产业适配性的核心技术支撑国家数字化转型。生成式可视媒体技术作为图形学与 AI 深度融合的前沿成果，突破了传统视觉内容生产对模板、规则和人力密集的依赖，具备“模型构建—内容生成—语义理解”全链条自驱的能力，成为实现跨模态内容建模、物理一致性生成与虚实融合表达的关键路径。

更为重要的是，该类技术对算力结构的适配性强，已成为推动我国 AI 基础设施国产化的重要应用牵引点。例如，基于扩散模型的图像与三维场景生成任务，天然适配图形处理器（graphics processing unit, GPU）/异构并行体系结构，带动芯片、操作系统、基础库等上下游生态同步演进。生成式视觉模型还为国产 AI 大模型提供了可落地的“具身智能”支持，使语言、视觉、动作等多模态输入具备统一表达语义的潜力，推动从感知智能向决策智能跃升。

此外，生成式可视媒体技术正在快速由研究突破

走向工程实践，在制造业、文创产业、国防安全、智慧城市、医疗教育等关键领域展现出强大生命力。在智能制造领域，生成式 AI 正逐步嵌入工业设计、虚拟仿真、产品定制等环节。通过与 CAD/CAE 系统深度结合，生成式模型能够实现结构几何的自动补全、制造工艺路径的智能规划，以及三维模型的逆向推理与优化设计，从而显著降低产品开发门槛，提升定制效率。例如，基于点云或网格的几何生成网络可实现从扫描数据直接重建参数化模型，大幅提高了高端制造装备在复杂环境下的适应性。在数字文创领域，生成式视觉内容极大地提升了影视、游戏、广告等内容的创作效率与形式创新能力。利用大规模预训练模型，可实现基于文本驱动的视频脚本分镜图生成、虚拟角色构建、游戏世界地图自动合成等高价值任务。同时，国产大模型已逐步具备文化风格迁移与传统艺术重构的能力，有望推动中国文化知识产权（intellectual property, IP）的智能再创造与全球传播。

生成式可视媒体技术不仅具备直接的应用价值，其背后的方法体系和算力需求还将对我国人工智能基础研究、芯片设计、开源生态等形成长远牵引力。一方面，面向产业需求推动国产视觉大模型的训练和优化，可加速形成“场景驱动—模型训练—系统落地”闭环式创新路径；另一方面，生成式内容还可作为高质量合成数据源，服务于算法鲁棒性提升和隐私保护等基础问题的研究。

在国家“东数西算”“人工智能+”和“数据要素市场化配置”政策推动下，生成式可视媒体将成为未来数字内容要素中的关键生产力。其在保障文化主权、提升城市治理韧性、支撑国家安全体系现代化等方面，展现出不可替代的战略地位。对生成式视觉基础设施的前瞻部署，将为我国从“视觉大国”迈向“视觉强国”提供系统支撑。

总结

在数字经济与智能化浪潮下，生成式可视媒体正加速重塑工业制造、文化创意、智慧城市等创新生态。其通过高效生成图形、图像、视频及沉浸式虚拟环境，不仅显著提升内容生产效率和个性化能力，还为智能制造、医疗健康、影视娱乐等行业开辟了新路径。例如，基于生成式 AI 的柔性产线优化、精准三维解剖模型重建和动态场景高效生成等应用展现出颠覆性潜力，有望成为推动我国数据要素建设和数字经济增长的核心引擎。然而，产业落地仍面临严峻挑战：在理论上亟须融合几何与物理属性的统一表达机制，突破视觉特征提取与多模态认知对齐瓶颈；在应用上需推动与工业级 CAD/CAE 等系统深度结合，实现设计—仿真—制造全流程智能化。同时，高质量数据匮乏、计算架构效率不足和生成内容可控性薄弱等问题制约了规模化应用。更为紧迫的是，全球竞争加剧，国际巨头加速布局，若我国未能实现关键技术自主突破，将面临技术依赖与产业主导权旁落的风险。

为此，应从国家战略层面统筹资源，构建“技术攻坚—生态协同—应用落地”的全链条创新体系：聚焦生成式可视媒体的十二大科学技术问题，加大基础理论研究投入；推动跨学科协作，建设开放共享的工业数据集与评估标准；强化国产化软硬件生态，突破芯片、框架等“卡脖子”环节；同时，完善伦理治理与版权保护机制，保障技术健康发展。通过凝聚产学研合力，抢占生成式可视媒体全球制高点，为我国数字经济高质量发展注入新动能。 ■

致谢：感谢清华大学胡事民院士、浙江大学鲍虎军教授、北京大学汪国平教授、上海交通大学马利庄教授、北京大学彭宇新教授、上海科技大学虞晶怡教授、复旦大学张新鹏教授、中国科学院软件研究所田丰研究员以及中国科学技术大学陈雪锦、网易范长杰、中国科学院计算技术研究所高林、南京大学过洁、中国科学院自动化所赫然、vivo 李博、清华大学刘烨斌、北京大学刘利斌、山东大学吕琳、南京大学王贝贝、北京航空航天大学王莉莉、华为王铭学、中南大学夏佳志、国防科技大学徐凯、浙江大学杨易、华为郑士胜、浙江大学周昆、大连理工大学杨鑫、南京邮电大学张肇轩、大连理工大学赵佳岳在这期秀湖会议筹备中提供的宝贵建议和观点。

Generative Visual Media in the Era of Foundation Models: Opportunities and Challenges—Insights from the 23rd CCF Beautiful Lake Seminar

Compilation: Xin Yang[1], Beibei Wang[2], Jie Guo[2], Jiazhi Xia[3], Lili Wang[4], Lin Lyu[5], Libin Liu[6], Xuejin Chen[7], Lin Gao[8], Ran He[9], Kai Xu[10], Kun Zhou[11]

1. Dalian University of Technology
2. Nanjing University
3. Central South University
4. Beihang University
5. Shandong University
6. Peking University
7. University of Science and Technology of China
8. Institute of Computing Technology, Chinese Academy of Sciences
9. Institute of Automation, Chinese Academy of Sciences
10. National University of Defense Technology
11. Zhejiang University

Abstract: Generative visual media, encompassing images, videos, 3D geometry, as well as virtual reality (VR) and augmented reality (AR), is becoming a transformative force in the digital economy. Driven by diffusion models, Transformer architectures, and large-scale multimodal pretraining, it is reshaping traditional content creation paradigms and enabling more realistic, controllable, and high-fidelity generation. This report summarizes the key insights of the 23rd CCF Beautiful Lake Seminar, which brought together experts nationwide to discuss the theoretical foundations, computational architectures, and interdisciplinary applications of generative visual media. The discussions focused on major challenges such as geometry-physics integrated representation, native 3D feature extraction, multimodal alignment and control, CAD/CAE system integration, and multisensory VR content generation. The symposium identified the core bottlenecks, challenges, and development pathways of generative visual media and reached a consensus on future actions. This report distills twelve key scientific and technological questions, outlines five categories of application scenarios, and highlights the role of generative visual media in advancing China's technological breakthroughs and industrial deployment, providing strategic guidance for its application in intelligent manufacturing, cultural industries, and national security.

Keywords: visual media computing; generative models; computer graphics; virtual reality; computer-aided design; controlled generation

摘　要: 生成式可视媒体涵盖图像、视频、三维几何以及虚拟现实（virtual reality, VR）、增强现实（augmented reality, AR）等媒体形式，正成为推动数字经济发展的变革性力量。在扩散模型、Transformer 架构与大规模多模态预训练技术的驱动下，生成式可视媒体重塑了传统内容创作范式，实现了更加真实、可控、高保真的内容生成。本报告总结了第二十三期 CCF 秀湖会议的核心观点，会议汇聚了来自全国的领域专家，围绕生成式可视媒体的理论基础、计算架构及其跨学科应用展开深入研讨。会议讨论聚焦于几何–物理一体化表示、原生三维特征提取、多模态控制对齐、计算机辅助设计（computer-aided design, CAD）/计算机辅助工程（computer-aided engineering, CAE）系统集成、多感官 VR 内容生成等关键技术挑战，并总结了“生成式可视媒体”的技术瓶颈、核心挑战与发展路径，最终达成行动共识。本报告通过凝练 12 项关键科学技术问题，总结生成式可视媒体的五类应用场景，并阐述生成式可视媒体技术对我国技术突破与产业落地的推动作用，为生成式可视媒体技术在我国智能制造、文化产业、国家安全等关键领域的部署提供了方向指引与技术路径。

关键词: 可视媒体计算；生成式模型；计算机图形学；虚拟现实；计算机辅助设计；可控生成

中图分类号: TP18

中文引用格式: 杨鑫, 王贝贝, 过洁, 等. 大模型时代下生成式可视媒体的机遇与挑战——第二十三期 CCF 秀湖会议报告 [J]. 计算, 2025, 1(6): 68−78.

英文引用格式: Xin Yang, Beibei Wang, Jie Guo, et al. Generative Visual Media in the Era of Foundation Models: Opportunities and Challenges—Insights from the 23rd CCF Beautiful Lake Seminar[J]. *Computing Magazine of the CCF*, 2025, 1(6): 68−78.

（本文责任编委：杨　易）