

视觉 AIGC 的文化学习边界：基于表征距离模型的理论解释

杨 洸¹ 余青莲^{2*}

1. 湖北美术学院, 湖北 武汉 430060

2. 武汉设计工程学院, 湖北 武汉 430060

摘要：视觉 AIGC 在文化图像生成领域已展现出较强的视觉模仿与风格迁移能力，但对于制度规范、价值观念等深层文化知识的学习仍存在明显局限。现有研究关注的重点在文化知识的组织与输入，而对文化知识为何难以被视觉模型学习缺乏系统解释。本文从文化知识学习机制视角出发，提出“表征距离 (Representation Distance)”概念，将其定义为文化知识转化为视觉表征所需经历的语义转换层级数量，并构建表征距离模型用于解释视觉 AIGC 的文化学习边界。进一步提出文化知识分层转译机制，引入文化中间表征层 (Intermediate Cultural Representation, ICR)，通过构建结构性转译与叙事性转译两种路径，实现高层文化知识向视觉语义的逐级映射，从而缩短知识与视觉表征之间的表征距离。本文尝试将视觉 AIGC 文化研究的关注重点由知识组织转向知识学习机制，为文化智能生成系统的构建提供新的理论视角与方法论参考。

关键词：视觉 AIGC；文化学习；表征距离；文化中间表征层；分层转译机制

DOI: 10.64649/yh.shygl.issn3105-0085.202606026

1 研究背景与研究问题

随着生成式人工智能技术的发展，视觉 AIGC 已广泛应用于艺术创作、文化遗产数字化及文化创意设计等领域。依托大规模图文数据训练，模型能够快速生成具有特定文化风格的视觉内容，在传统文化数字传播与创新表达方面展现出巨大潜力。现有生成结果往往能够准确模仿文化符号与视觉风格，却难以表达其背后的文化内涵与生成逻辑，呈现出“形式丰富而语义贫乏”的特征。目前的相关研究主要从三个方面展开探索：一是通过文化元素提取与领域数据集构建增强模型的文化特征学习；如吴晶晶等通过文化感性因子提取、领域模型训练与用户需求映射，提出一种融合感性工学的非遗文创智能设计方法，实现用户感性驱动的文创产品生成^[1]。陈永等提出一种基于布朗桥扩散模型的古壁画修复方法，通过有界噪声控制、掩膜感知上下文建模与逆向迭代重建，实现对壁画全局结构与局部细节的协同修复^[2]。

二是利用微调训练提升模型对特定文化风格的生成能力。如：宋玉等以齐白石先生的名画“蟹”系列为例，利用显著性检测技术提取其中的视觉信息，基于扩散模型，训练出能实现传统水墨画作特征的 AIGC 文创模型，并将生成结果应用于文化产品及创意衍生品的设计中^[3]。侯云鹏等提出一种基于 LoRA 微调与 Stable Diffusion 结合生成的楚漆器数字化传承方法，通过构建专用数据集和风格模型训练，实现楚漆器艺术特征的智能生成与创新应用^[4]。

三是借助知识图谱、语义标注等知识增强

方法提高文化信息输入密度。如：侯西龙等通过构建青铜器知识本体模型，规范描述多维度的青铜器知识体系，并使用“人工+大语言模型”等方式增强图像语义标注，实现青铜器风格图像的智能生成与创作^[5]。王伶俐等通过 AIGC 驱动的多模态知识库的构建、知识调取与融合机制、知识引导的可控生成等方法，有效整合图像、文本等多源数据，提升了文化遗产图像修复过程中风格一致性与语义连贯性^[6]。

这些方式在丰富文化知识供给方面取得了积极进展，但其基本思路均建立在“增加知识输入即可提升文化表达能力”的假设之上。然而，大量实践表明，即使持续增加训练数据与知识信息，模型对于制度规范、价值观念等高层文化知识的表达能力仍然有限。这说明文化学习能力并不完全取决于知识规模，文化知识本身可能存在差异化的学习边界，部分知识则需要经过复杂的语义转换才能获得视觉表达。因此，高层文化知识在进入视觉模型的过程中往往面临更严重的信息损失。针对上述问题，本文从文化知识学习机制视角出发，探讨视觉 AIGC 的文化学习边界及其形成原因。构建文化知识分层转译机制，通过建立文化知识向视觉表征的中介转换路径，降低高层文化知识在学习过程中的信息损失。

2 视觉 AIGC 的文化学习边界

2.1 视觉 AIGC 的知识学习机制

视觉 AIGC 的生成能力来源于对大规模图

文数据统计规律的学习。无论是文本编码器建立的语义关联,还是扩散模型学习的视觉分布,其本质都是通过视觉表征与语义表征之间的对应关系完成知识获取。因此,模型能够学习的前提在于相关知识能够稳定地转化为视觉信息。对于文化知识而言,模型并非直接学习文化本身,而是学习文化在视觉层面的表达结果。文化知识能否形成稳定、明确的视觉对应关系,将直接影响其进入模型的难易程度。

2.2 文化知识的视觉可表征差异

文化知识并不是均质结构,不同类型知识与视觉表征之间存在显著差异。形态知识主要表现为纹样、色彩、器物造型等视觉形态,本身即可作为图像内容存在;工艺知识虽然属于过程性知识,但往往能够通过材料特征、制作痕迹和形态结果获得部分视觉表达;制度知识则更多体现为行为规范、身份秩序和组织关系,需要借助特定活动场景才能获得可视化呈现;观念知识涉及价值认知、精神信仰与文化意识,其表达通常依赖复杂的社会实践与符号系统。不同文化知识进入视觉系统所需经历的转换过程并不相同。部分知识能够直接形成视觉对象,部分知识则需要经过多层语义转换才能获得视觉表达。

2.3 表征距离与文化学习边界

基于此,本文提出“表征距离(Representation Distance)”概念。表征距离并非指知识本身的抽象程度,而是指文化知识转化为视觉表征所需经历的语义转换层级数量。转换层级越少,知识越容易形成稳定的图文对应关系;转换层级越多,知识在转换过程中发生信息损失的可能性越高。例如,形态知识能够直接对应视觉对象,其表征距离最短;工艺知识需要通过工艺结果获得视觉呈现;制度知识需要进一步借助行为活动与场景关系完成表达;观念知识则往往需要经过行为、场景乃至符号系统等多个中介环节才能形成视觉表征。因此,文化知识的学习边界并非简单由模型能力决定,而是受到知识与视觉表征之间距离结构的影响。表征距离越长,文化知识越难通过现有视觉学习机制获得完整表达。

现有文化增强研究普遍将提升路径集中于知识规模扩充与知识结构优化,其核心目标是向模型输入更多文化信息。然而,当文化知识无法有效转化为视觉表征时,知识输入的增加并不会自动转化为学习能力的提升。因此,视觉AIGC面临的关键问题并非文化知识不足,而是缺乏将高层文化知识转化为视觉语义的有效机制。如何缩短文化知识 with 视觉表征之间的表征距离,成为突破文化学习边界的核心问题。

3 文化知识表征距离模型及其转译机制

3.1 表征距离模型构建

不同类型文化知识在视觉AIGC中的可学习性存在显著差异。这种差异并非源于知识内容本身,而是知识转化为视觉表征过程中所经历的语义转换过程。基于此,本文提出表征距离(Representation Distance)模型。其并非知识本身的抽象程度,而是文化知识转化为视觉表征所需经历的语义转换层级数量。当某类知识能够直接形成视觉对象时,其表征距离较短;当知识需要借助多个中间环节才能获得视觉表达时,其表征距离则相应增加。表征距离越长,知识在转换过程中发生语义损耗的风险越高,被视觉模型有效学习的难度也越大。表征距离可以被理解为文化知识进入视觉生成系统的学习成本。

3.2 表征距离的量化描述

为进一步描述不同文化知识的学习难度,本文尝试对表征距离进行结构化表达。设文化知识到视觉表征之间需要经历的语义转换层级数为 N , RD 表示表征距离(Representation Distance),表征距离可表示为: $RD = N$ 。例如,纹样、色彩等形态知识能够直接形成视觉对象,其转换层级数较少;工艺知识需要经过工艺过程与工艺结果两个阶段;制度知识则需要通过行为规范、仪式活动与场景组织获得表达。观念知识通常还需要经历价值认知、行为模式与场景叙事等多个中介环节。由此可见,不同文化知识对应不同长度的语义转换链,高层文化知识往往具有更长的表征距离。

然而,仅以转换层级数衡量学习难度仍不充分。由于同一文化知识可能对应多种视觉表达路径,因此表征距离还受到语义分叉程度的影响。设知识在转译过程中产生的平均语义分支数,即转译过程中语义表达的不确定程度为 B ,则表征距离可进一步表示为: $RD = N \times B$ 。转换层级越多,语义分叉越复杂,知识进入视觉模型的难度也越高。

3.3 文化知识的表征距离结构

依据表征距离模型,不同文化知识可形成递进式距离结构。形态知识与视觉表征之间具有直接对应关系,其表征距离最短;工艺知识需要通过技术过程获得视觉表达,距离有所增加;制度知识需要借助行为活动与空间组织实现表达;观念知识则需要经过意义生成、行为实践与场景叙事等多个环节才能形成视觉表征,因此具有最长的表征距离。这种差异说明,高层文化知识并非天然无法学习,而是在知识向视觉转化过程中面临更长的语义转换链与更高的不确定性。因此,文化学习边界本质上体现

为表征距离不断增加所产生的信息衰减现象。

3.4 基于表征距离的分层转译机制

既然文化学习边界源于表征距离过长,那么提升文化学习能力的关键就在于缩短知识与视觉表征之间的距离。基于这一思路,本文提出文化知识分层转译机制。其核心目标是在文化与视觉表征之间建立稳定的文化中间表征层(Intermediate Cultural Representation, ICR),将原本需要跨越多个转换环节的知识转化过程拆解为连续且可学习的语义映射过程。ICR并非新的知识层级,而是一种语义转译结构,形成“文化知识→ICR→视觉表征”的转译路径。其作用是将高层文化知识转化为具有视觉可表征特征的中间语义单元,从而降低知识进入视觉系统时的转换成本。

3.5 两种高层文化知识的转译模式

对于高层文化知识而言,不同知识类型具有不同的视觉转译逻辑。制度知识的核心特征在于组织关系与秩序建构,其视觉表达主要体现在对象之间的结构关系。因此,制度知识更适合采用结构性转译模式(Structure-based Representation):“制度知识→仪式行为→场景结构→视觉表征”。关注重点是视觉对象如何组织、排列与关联。

观念知识的核心特征在于意义生成与价值表达,其视觉呈现通常依赖行为过程与情节发展。因此,观念知识更适合采用叙事性转译模式(Narrative-based Representation):“观念知识→行为模式→场景叙事→视觉表征”。

参考文献:

- [1] 吴晶晶,赵慧雯.基于感性工学与AIGC的泉州提线木偶文创设计转译研究[J].包装工程,2025,46(16):308-319.DOI:10.19554/j.cnki.1001-3563.2025.16.025.
- [2] 陈永,范志欣,张世龙,等.布朗桥动力学控制的古壁画AIGC算法[J/OL].控制理论与应用,1-7[2026-06-14].https://link.cnki.net/urlid/44.1240.TP.20260304.2112.008.
- [3] 宋玉,钱晓松,彭凌钰,等.AIGC中式美学文创设计研究[J].包装工程,2023,44(24):1-8+33.DOI:10.19554/j.cnki.1001-3563.2023.24.001.
- [4] 侯云鹏,彭涵,刘育晖.基于LoRA模型的非遗数字化传承:以楚漆器为例[J].设计艺术研究,2024,14(1):14-18.
- [5] 侯西龙,王雨鑫,臧君涵,等.基于AIGC的文化遗产图像智能生成方法研究——以青铜器为例[J].图书情报工作,2026,70(5):56-68.DOI:10.13266/j.issn.0252-3116.2026.05.005.
- [6] 王伶俐,方明珠,郭美星,等.AIGC多模态知识驱动的文化遗产图像修复发展现状[J].包装工程,2026,47(4):1-18.DOI:10.19554/j.cnki.1001-3563.2026.04.001.

作者简介: 杨洸(1983.12—),男,郑州,汉,本科,中级实验师,研究方向:人工智能与金属工艺。

通讯作者: 余青莲(1984.07—),女,武汉,汉,本科,教授,研究方向:工艺美术。

项目信息: 2023年度湖北省教育厅科学研究计划指导性项目:B2023383《增材制造技术的楚漆器文化创意产品开发研究》。

关注重点是视觉场景中发生什么以及为何发生。两种转译模式共同构成高层文化知识进入视觉生成系统的主要路径,也构成分层转译机制的核心内容。

4 表征距离模型的理论意义与应用启示

在表征距离框架下,训练数据不仅需要记录文化对象,还需要记录知识向视觉表达转化过程中所经历的中间语义环节。因此,未来文化训练数据的组织重心应由静态知识描述逐步转向动态转译过程描述,通过保留知识转译链条中的关键语义节点,为视觉模型提供更加完整的文化表达路径,为文化训练数据构建提供了新的理论依据。

本文从理论层面对视觉AIGC的文化学习边界进行了讨论,并提出表征距离模型及其分层转译机制。由于研究重点在于理论建构,尚未对不同文化知识的表征距离进行具体测度,也未结合实际模型开展实验验证。未来研究可围绕三个方向展开:一是进一步建立表征距离的量化评价方法,探索不同文化知识类型的距离测度标准;二是研究文化中间表征层的构建规则与自动生成机制;三是结合扩散模型、多模态大模型与知识增强系统,验证分层转译机制对文化生成质量的影响。表征距离模型揭示了文化知识进入视觉生成系统过程中存在的结构性障碍,为理解视觉AIGC的文化学习机制提供了新的理论视角,也为未来文化智能生成系统的构建提供了方法论参考。