

# 谄媚的图像

## ——技术哲学视角下计算机生成图像批判

○ 耿游子民

**摘要：**今天，计算机生成图像（CGI）业已建立巨大的技术乌托邦幻觉，祛除幻觉成为正确认识 CGI 的前提。本文通过对电子游戏图像进行媒介物质性分析，总结出三类视知觉欺骗——实在、象征与想象界欺骗。而后分析人工智能介入 CGI 后的技术特征：潜空间与模式幻觉，指出 CGI 内在机制从“图像的否定性规训”走向“现象学还原的断裂”。最后，结合韩炳哲和巴迪欧的相关理论指出当下 CGI 从欺骗图像走向谄媚图像，其所承诺的世界模拟只是一种对人类中心主义的谄媚。

**关键词：**电子游戏图像 视知觉欺骗 技术哲学 平滑世界 事件

### 引言：你也渴望模拟世界？

“作为视频生成模型的世界模拟器（Video generation models as world simulators）”<sup>[1]</sup>，这是 OpenAI 公司旗下人工智能视频生成模型 Sora 技术报告的标题。模拟（Simulate）来自拉丁语“simulare”，最初意为模仿或假装。而随着 20 世纪计算机技术发展，模拟一词从日常生活中脱离，明确指向某种对现实世界部分乃至全部现象的仿真，继而出现了对应的技术物概念——模拟器（Simulator）。其实模拟器这一概念并不新鲜，早在 1981 年，希拉里·普特南（Hilary Putnam）就在其著作《理性、真理和历史》一书中提出了“缸中之脑”<sup>[2]</sup>这一思想实验<sup>[3]</sup>，也成为日后《黑客帝国》（*The Matrix*）等一系列佳作的灵感来源。而后，斯蒂芬·哈特曼（Stephan Hartmann, 1996）从过程哲学出发讨论模拟的巨大潜能。<sup>[4]</sup>尼克·博斯克洛姆（Nick Bostrom, 2003）则尝试从认识论层面对“我们是否生活在计算机模拟之中”这一问题给出肯定回答。<sup>[5]</sup>

今天，视觉承载了人类最大的知觉来源，其中绝大部分又以图像呈现。通过图像进行“世界模拟”似乎是

耿游子民，北京大学艺术学院 2023 级博士研究生。

一条颇为合理的路径。更为有趣的是，虽然至今没有公测（本文写作阶段，即 2024 年 12 月 12 日 Sora 公测）但有关于 Sora 的讨论早已层出不穷。向如此华丽并伴着巨大野心的媒介投入热情无可厚非，但对于技术的审慎与警惕恰恰是人文学者所必备的眼光。贝托尔特·布莱希特（Bertholt Brecht）在《作为交流工具的无线电》（*The radio as an apparatus of communication*）一文中曾告诫我们：

我们经常高估新的媒体形式，因为它所承载的东西有着乌托邦式的期望，这些对广播有很高评价的人之所以拥有它，是因为他们在其中看到了可以发明“某些东西”的东西。他们会自认为是正当的。<sup>[6]</sup>

的确，对于技术的过分信任与期待往往导致“伊卡洛斯的悲剧”。黑箱的不断叠加让我们只需关心技术的便利，却无视了黑箱之下凛冽的真实，对于新兴技术的局限性讨论迫在眉睫，基特勒曾在实体之夜中阐发人类面对技术黑箱的态度：“留声机在唱片声槽间的跳跃，一个新的电话号码的拨出或者两个电影图像间的剪切处都横贯了黑胶、黄铜电缆或者胶片观赏系统欺骗眼睛中那持续的数据流。”<sup>[7]</sup>既然 Sora 至今难以触及，作为其上层范畴的计算机生成图像便可纳入讨论，沿着媒介物质性视角敞开黑箱，对其加以分析将成为祛魅“世界模拟器”的第一步。

## 一、从 CGI 到电子游戏：技术耦合下的象征失序

计算机生成图像（Computer Generated Imagery）由来已久，早在 20 世纪 50 年代的旋风计算机（Whirlwind）上便出现了计算机生成的抽象图像，而生成非抽象图像最早可追溯至前苏联大型机 BESM-4。1968 年，康斯坦丁诺夫等人使用一组方程定义了一只猫的轮廓。计算机通过计算该组方程，将每一个结果以坐标点的形式绘制在碳带（一种早期打印材料）上再由胶片翻拍后播放，获得了最早的 CGI。<sup>[8]</sup>1982 年，Muse Software 为家用电脑 Apple II 开发的动作冒险游戏《德军总部》（*Wolfenstein*, 1982）代表着作为电子游戏图像的 CGI 的彻底独立，而作为本地实时演算开端的《死亡陷阱 II》（*Will The Death Trap II*, 1985）则包含了 CGI 成为“视觉世界模拟器”的初步可能。克劳斯·皮亚斯（Claus Pias）在《电子游戏世界》中叙述道：“用户不需要用手做圆周运动，而是通过输入一个中心点和圆周点，剩下的由计算机来完成。”<sup>[9]</sup>伊恩·博格斯特也表示：“编写代码来执行规则继而产生表达物，而非直接创造表达物本身”<sup>[10]</sup>。二者所指向的便是 CGI 同其他图像记录或绘制的本质不同，这是一种符号自生成系统。自然，以生成为本质属性的电子游戏图像成为媒介物质性层面 CGI 的直接代表，其生成与记录过程极端的隐匿于技术物黑箱之中。但另一方面，这些图像最终依然要面对人类的知觉评价。一言以蔽之这是一种“予人”的图像。正是 CGI 这种予人的本质属性暗含了 CGI 作为视知觉欺骗的诡诈之处，一如基特勒

在《留声机电影打字机》中挪用拉康的三界理论，我们同样可以将其对于人类的视知觉欺骗分为三类加以讨论。

### （一）实在界欺骗，离散的知觉妥协

首先需要明确，现实世界与电子游戏以不同的方式呈现“连续”。在电子游戏中，“帧”（Frame）是其时间测量的最小单位。游戏运行时，任意物体的空间坐标往往以帧为单位进行刷新，而物体边缘的空间坐标发生交叉往往预示着二者发生了碰撞。现实世界的碰撞展现为物体相切后



图 1 《平行 III》（*Parallel III*），哈伦·法罗基（2014）



图2 《星球大战:陨落的武士团》(Star Wars Jedi: Fallen Order, 2019), 游戏截图

彼此发生形变的“瞬间”，而电子游戏的“帧”作为离散时刻，实则无法刻画碰撞这一简单的物理现象：快速移动的物体A与B前一帧彼此相离，但下一帧彼此空间位置便可能相互叠加，在视觉上呈现为A与B物体的模型彼此相交也就是所谓的“穿模”的错误状态。为了解决这个问题，通常情况下物体间的碰撞检测通过无法看到的碰撞体(Collider)实现，这些碰撞体通常比可知觉的模型稍大，扩大比例往往与物体在空间中的运动速度成正比，以避免穿模的情况发生。哈伦·法罗基(Harun Farocki)的《平行III》(Parallel III, 2014)中有大量基于这一现象的创作(图1)。

同样，电子游戏中以“像素”作为其空间识别的最小单位，当物体或模型远离观察者时，其占据的像素面积不断缩小，呈现为较小的分辨率。为了配合这一特性，计算机图形学发展出了LOD(Level of Detail)技术——即多细节层次。指根据物体模型在显示环境中所处的位置和重要度，决定物体渲染的资源分配，通过降低非重要物体、较远物体的多边形面数和材质细节度从而获得较高的渲染效率。例如图2便展示了一个典型的LOD错误，图中本应呈现为高多边形数量与高分辨率材质贴图的岩石与地面错误渲染为了在较远距离才会呈现的状态。诸如此类的欺骗方式还有很多，如只渲染玩家视角内呈现画面的视锥剔除(Frustum Culling)、光栅化(Rasterization)、动态场景加载等等。

因此，电子游戏图像所存在的“世界”同现实有着本质差异，这种差异体现且不仅体现在时空计量上。从本体论角度来看，作为构成这一世界的基底，电子游戏中离散的“帧”与“像素”构成了对实在界的片段化模拟，以一种非连续的有限性隐藏了其实在界褶皱。<sup>[11]</sup>所谓的“技术迭代”——如提高帧率或提高像素密度，只是将褶皱以一种更为隐蔽的方式均匀地分布到对实在界的想象之中，它们无法真正还原世界的连续性。这种分布并没有从本体论层面解决其与现实的根本区别——根据热力学第二定律，能量的利用效率不可能达到完美，而数字图像模拟的计算过程同样受到物理资源(如能量、算力等)的限制。因此，无论帧率或像素密度如何提高，都无法从根本上突破技术的离散性。这种限制不仅源于计算复杂度，也源于现实世界的连续性与数字媒介的离散性之间的本体论矛盾。所以模拟的本质只能是计算机算力与图像真实感之间的博弈，即在有限资源下最大化图像的“视觉近似性”。而最终用以评价模拟水平的同样并非现实世界，而是视觉本身。从这一层面出发，实在界欺骗表现为对现实物理现象的知觉妥协。设计者利用感知漏洞与局限性调整图像渲染的方式，使得玩家感知到一个连续且一致的虚拟世界，每一次提高帧率或增加像素密度都仅仅是让结构的离散更加难以察觉，并不能从根本上改变数字图像世界与现实世界间的本体论差异。

## (二) 想象界欺骗:幻象的视觉补全

相比于知觉无法直接接触的实在界，想象界作为中介、弥合、填充实在界与象征界裂隙的领域，依赖于玩家建立指向于自身的知觉模式，而CGI对于想象界的欺骗同样围绕视觉有限性展开，利用玩家的主观感知来构建拟仿现实呈现的虚拟体验。在这个过程中，想象界的符号体系以一种同现实无关甚至反对现实的状态呈现图像。

一个经典案例是《圣域2:堕落天使》(Sacred 2, 2008)



图3 《圣域2:堕落天使》模组《冰与血》(2008), 游戏截图



图4 左侧树木图像正常显示状态,右侧为强行变化视角而产生的错误图像

中的《冰与血》(Ice & Blood)模组。尽管该模组并未使用实时光线追踪(Ray Tracing)或立方体贴图(cube mapping)等技术,却实现了高质量的水晶反射效果。这一效果的关键在于“使用最后渲染的帧作为反射源”(图3)。这个过程可以总结为:首先,选中一个需要反射

的像素并调整其法线方向,决定反射的来源。然后,查看上一帧已经渲染好的画面,找到调整后法线指向的位置,最后将那个位置的像素绘制到需要反射的像素上。这样,就可以利用前一帧的图像数据来模拟反射效果,而不需要实时光线追踪。

因此,这个反射总是显示游戏“前一刻”的画面,但鉴于帧与帧间隔时间如此之短,这种延迟在视觉上难以精确识别,往往会被人眼归纳为合理的连续反馈,这正是以一种无关现实的状态呈现的图像。而《暗黑破坏神3》(Diablo 3, 2012)中的树木也有着同样的处理,美术设计师将特定角度的树木纹理绘制在倾斜的平面上,虽然树木只能从一个角度观看,但大大减少了多边形数量与贴图尺寸,以至于可以呈现十分丰富的细节。而《暗黑破坏神3》作为一个2.5D游戏,其无法进行镜头旋转的特性也决定了这些树木只能以固定方向观看,不会出现旋转视角后产生的错误画面(图4)。另一个案例则来自《古墓丽影9》(Tomb Raider 9, 2013),游戏中出现的燃烧火把使用了典型的可循环预渲染纹理图像序列实现。通常情况下这样的欺骗效果十分完美,但假如劳拉进行大幅度动作,火焰贴图将被迫发生大幅度的错切与形变,暴露出其图像序列而非真实物理模拟的本质(图5)。

当然,类似的案例不计其数,尤其是涉及到计算机生成三维图像这一技术热点。在任天堂公司推出的便携掌上游戏机GBA上就出现过通过对二维图像的实时缩放、旋转和错切实时模拟三维效果的完整技术体系“Model 7”<sup>[12]</sup>,其次诸如法线贴图(Normal Mapping)、视差贴图(Parallax Mapping)等技术也被用来制造特定的视觉假象。

总而言之,CGI在想象界所构建的假象并不仅仅是“技术限制下的产物”,更是主观感知主动迎合技术的结果。想象界作为符号中介,在人类的知觉系统中起到弥合实在与象征界的作用。CGI通过精妙的设计,利用玩家的视知觉模式局限生成了一种介于真实和罔顾真实之间的视觉体验。这种拟仿现实并不依赖对客观物理世界的精确复现,而是依赖玩家根据自身经验与认知模式主动补全图像中的不完美之处。因此,想象界欺骗本质上是对人类感知期待的回应——CGI为玩家设计了一个有限结构,玩家的知觉则通过感知自动弥补细节空白。这一机制是电子游戏图像欺骗的核心手段之一。数字系统并不试图还原现实,而是通过引导玩家的知觉,使其相信自己看到了与现实一致的场景。这种欺骗并非基于对现实的模拟,而是基于玩家视觉模式中的“补全”,一种视觉的逻辑性延迟。欺骗成为玩家缝合实在与象征界的“规则”,在感知过程中,玩家忽视视觉中的矛盾,将其整合为连续且合理的视觉经验。

这种视觉规训不仅在电子游戏中被广泛运用,在任何CGI系统中玩家的知觉习惯都通过这种欺骗不断被重塑。由此可见,玩家实际上参与了视觉幻象的构建过程,成为了技术骗局的一部分。在技术引导下,玩家主动为视觉模式填补空白,成为了技术所构建世界的一环。这并非简单的感知欺骗,而是通过图像结构化,促使玩家的认知体系为技术服务,构成了“人机共谋”。这意味着,在技术规训下,玩家的自我感知被系统引导并限制,成为技术物个体化过程中的一部分。



图5 《古墓丽影9》(Tomb Raider 9, 2013),游戏截图



图6 《超级马里奥兄弟》(Super Mario Bros, 1985), 游戏截图

### (三) 象征界欺骗:符号的逻辑规训

象征界作为人类通过结构化符号所建立的社会秩序,规定了个体如何与世界互动、表达欲望、理解自我与他人的基本逻辑。一如基特勒将打字机指认为“象征界”媒介,电子游戏图像于象征界同样规定了自治的符号系统,通过这一符号体系的规训使玩家在视觉上承认游戏中的现象,哪怕这些现象以一种与“现实无关”的状态出现。针对象征界欺骗的一个经典案例便是在《超级马里奥兄弟》(Super Mario Bros, 1985)游戏中,草坪和白云使用了相同的图像,唯一的区别是通过颜色改变区分了画面内容(图6)。

同样的例子包括《1943》中的阴影。在游戏中,阴影不过是与战斗精灵图(Sprite)<sup>[13]</sup>一致的黑色剪影,与真实世界中阴影的产生方式毫无关系(图7)。在这两个例子中,游戏图像首先特异地建立了符号系统内任意的所指-能指关系,继而在这种简单的关系模式上借由视觉特征变化让玩家承认其知觉到了不同对象。显然,这种特异性的符号映射除了上述仍部分拟仿现实存在的例子外,更多的是彻底抛弃模拟的存在。如《我的世界》(Minecraft, 2010)通过完全由立方体构成的世界建立了一个与现实对应符号映射系统。尽管现实中的物体有着丰富的形态和复杂结构,但在此处的电子游戏图像中,所有的物体——无论是树木、山脉,还是生物都被统一为规则的立方体形态。玩家通过上述的符号视觉系统,接受并内化了与现实大相径庭的象征秩序与图像逻辑,在这个程度上与其说是一种欺骗,不如说是电子游戏通过符号建立的象征秩序,成功地将玩家的视觉感知引导至一种与现实无关的例外世界中,即使这些图像没有现实的支撑,玩家同样可以在这一象征秩序内理解、认同并参与互动,实现了某种意义上的模拟。

当然,象征界的欺骗依赖于象征秩序的建立与运行。而当这一秩序因外在循环被破坏时,问题也随之诞生,这种欺骗背后的凛冽便以洞穿实在界与想象界的姿态降临。正如《我的世界》在丢失材质贴图后,树木、山脉、生物被紫黑相间的方块取代,失去业已建立的象征秩序坍塌为无法把握的陌异图像(图8)<sup>[14]</sup>。从电子游戏图像的发生学层面上看,这种象征界欺骗内蕴在电子游戏图像的生成之中,只有当玩家全情投入且毫无反思的沉浸才能确保这套机制的正常运行。

从上述例子的分析中不难看出,实在界欺骗强调数字图像世界与现实世界的本体论差异;想象界欺骗关注人的主体性视角如何动摇“模拟”概念本身的本体论地位;而象征界欺骗则揭示了数字图像世界运行的底层逻辑。学者李洋曾对电子游戏下过一个定义——“一种工具性体验”<sup>[15]</sup>。这一定义给予了电子游戏图像批判一种有趣的本体论结构。在海德格尔那里,在手性与上手性作为工具的两个面向被平等的展开,而在电子游戏图像所构成的工具性体验这里,“在手性”作为模拟的对立面被主动地遮蔽,甚至可以说被玩家—游戏构成的行动者网络结构性地抛弃了,一旦有任何揭示“在手性”的可能——哪怕是某种潜在的倾向而非揭示本身,都会回溯性地瓦解这种工具性体验。注意,这是一种回溯性的瓦解,这意味着在手性在这一结构的建立中不能也无法留下位置,电子游戏图像作为一种工具性体验的前提是对在手性的绝对抛弃。而这正是对计算机生成图像批判的主要支点,不过在展开讨论之前,我们还需要引入新的技术变量——人工智能。

## 二、人工智能图像:断裂的数字神话



与直觉相反,人工智能生成图像(Artificial Intell- 图7 《1943》中的战斗机阴影



图 8 《我的世界》(Minecraft)

igence Generated Imagery, AIGI)并非一个晚近概念,今天的 Sora 也被大量从业者指认为使用了虚幻引擎 5 (Unreal Engine 5) 加以训练。可以说 AIGI 的历史与电子游戏图像,换言之与 CGI 的历史相差无几。20 世纪 60 年代,哈罗德·科恩(Harold Cohen)就尝试开发了一套用于绘画的计算机程序:亚伦(AARON)。<sup>[16]</sup>与今天围绕机器学习和神经网络建立的人工智能不同,彼时的亚伦仍是基于传统架构的计算机程序,不过借助复杂的判断逻辑与负反馈调解机制以及长期迭代,亚伦实现了比同期绘图程序更加复杂且多变的功能,从数量上更接近彼时对人工智能的幻想,但其生成的图像在本体论层面仍被归纳为一种和电子游戏图像同质的技术耦合图像。而后,卷积神经网络(CNN)、深度神经网络(DNN)、对抗式生成网络(GAN)先后涌现,引发了自十年前开始的人工智能新浪潮。总体上,人工智能技术的发展均与图像的生成直接相关,其中同样蕴含着“欺骗”这一概念,受篇幅所限不再赘述。此处着重考察与 CGI 有明显本体论差异,内蕴新发生学机制的扩散模型(Diffusion Model)。<sup>[17]</sup>

当下,包括 Stable Diffusion、Midjourney、DALL·E2 等文生图应用,以及 Sora、Runway、可灵等文生视频等应用,几乎都基于扩散模型(Diffusion Model)实现其功能。简单来讲,扩散模型分为两个部分,扩散(Diffusion)与去噪(Fusion),由于这个过程与后文现象学还原的断裂直接相关,故有必要详细演绎。想象有一幅清晰的关于狗的位图<sup>[18]</sup>,尝试给这个图像逐渐增加噪声——表现为图像中出现随机颜色的像素点,整个画面越来越模糊,直到最后完全看不清内容,画面变成了绝对随机,均匀的噪点图像,这一过程称为扩散。接下来模型要做的就是相反过程——从一张完全均匀模糊的图像开始一点点去除噪声,逐渐把模糊的图像还原为原来的清晰图像。也就是去噪(fusion)。在第二个过程中,模型需要学会如何“去噪”,训练时,模型将被提供大量真实图像,程序先将它们变模糊,再训练模型将噪声去除。随后,CLIP 模型<sup>[19]</sup>参与到流程中,引导去噪过程还原“文本所映射的模式”——换言之文本所指向的图像。不难发现,这个过程可以被分解为构成图像的每个像素的具体运算,但对于一张分辨率为 512 像素的图片,其在 Stable Diffusion 里需要被表示为一个  $512 \times 512 \times 3 = 786432$  维的向量,处理如此高维的向量需要较大的算力,且直接计算是低效的。为此,被称为变分自编码器(Variational AutoEncoder, VAE)的生成模型被引入到扩散进程中。VAE 能够将高维向量(如图像)编码到一个低维表示——即潜空间(Latent Space)中,且它引入了概率推断(Probabilistic Inference),通过对潜空间建模,生成的不再是映射而是概率分布。这使得 VAE 得以从潜空间中采样以生成更“逼真”的新数据样本。以上文提到的 512 像素图片为例,其可以被压缩为更小的  $64 \times 64$  向量模型,使得后续计算得以在更低维度的潜在空间中进行,大幅降低算力消耗,使得扩散模型真正具备实用价值。

恰恰在潜空间这里,可供现象学还原<sup>[20]</sup>的过程发生了根本性断裂。换言之,技术物所内蕴的因果系统被打破了。这点在西蒙东的《论技术物的存在模式》中有着充分讨论,西蒙东反复强调了技术物所内蕴的因果系统:“一个技术组合内的技术物的个性化原则是缔合环境中循环因果关系的子集合的原则”。这种因果关系即现象学还原的过程,直接的对比便是上文所述的扩散模型在未压缩的高维空间计算时,任一像素都能直接对应扩散模型中的一个向量,我们可以从直观经验中把握这个向量几何运算前后所发生的变化——也就是因果,推广到每一个向量,我们应然地把握着 AI 生成图像。这也是西蒙东强调的因果系统的锯齿状关系:

因此，在技术物的进化过程中，我们看到了因果关系的过渡，这种因果关系从先前的组合过渡到后继的元素。这些元素被整合到个体中，改变了后者的特征，容许技术因果关系再次从元素水平上升到个体水平，然后从个体水平上升到组合水平。在一个新的周期中，技术因果关系会下降到元素的制造过程，再到新的个体，然后是组合。因此，这样一条因果关系线不是直线而是锯齿状的，同一现实以元素的形式存在，然后是个体，最后是组合。<sup>[21]</sup>

然而，正如上文所说，潜空间是通过输入图像的高维特征进行降维和编码后形成的抽象表示，是一个通过训练数据建立的抽象空间，AI在其中操作图像的生成、组合和变化并非与像素一一对应——而是操作某种概率特征。潜空间并不是基于人类主观感知建构的，它基于数据与概率模式的动态过程而非结果，它的存在与人类经验世界没有关系。这种差异不仅导致了现象学还原的断裂，更动摇了人工智能生成图像的本体论地位，尽管它们“看似”真实，却缺乏人类体验中的意涵或现象学意义。详细来说，这种断裂至少发生在三个层面。首先是感知结构的断裂，扩散模型基于数据模式与概率统计生成图像，排除了人在图像世界中的经验，丧失了与世界的知觉交换。其次是基于数据局部限制的整体性断裂，扩散模型生成图像往往在局部合理但难以把握整体一致性，导致感知上的异常。最后，与CGI一样，潜空间生成图像同样是无时间的。虽然当下AI已经可以生成逼真的动态图像或视频片段，但这些图像并不存在通过时间积累的真实经验支撑。虽然Sora可以生成树林中的树叶飘动场景，但它并不存在对风力、重力、摩擦力等物理性质历时性感知，每一帧都是去矢量图像<sup>[22]</sup>，换句话说，AI生成的图像不存在“时间”。

除了潜空间产生的现象学还原断裂，AIGI难以规避的另一特性便是“模式幻觉(Algorithm Pareidolia)”。即由于训练数据或算法的局限，AI在随机过程中生成了貌似真实物体但实则虚假或存在二义性的视觉模式；或是AI过度依赖数据中的模式——过拟合，导致在没有上下文需求时也生成了某种重复结构或伪纹理，这一点在Deepdream<sup>[23]</sup>上得到了充分展示。无独有偶，人类自身也存在着类似的现象。例如经典的鸭兔图，其体现的并非某种视错觉，而是内蕴于人类模式识别的局限性，这种局限性的一大表征便是当图像恰好位于模式识别的模糊区域时产生的二义乃至多义性。

最后，不论是潜空间断裂还是模式幻觉，都指向着AI生成图像中的“平滑”现象。平滑指的是AI在图像生成过程中，通过去除噪音和模糊边界，生成柔和而连续的图像。尽管这可以提高图像的整体性并减少突兀感，但容易抹去现实中的粗糙感与微小细节，结果是图像在视觉上显得过于光滑反而不自然。究其原因，平滑现象与人工智能生成图像的底层算法密切相关，尤其是大量算法所依赖的概率统计模型。AI生成图像时，通常通过大规模数据训练和概率分布来生成视觉结果，而这些底层算法为了追求图像的连贯性和一致性，往往会倾向于生成“平均”的图像。这种平均化结果直接影响了图像的细节呈现，使得图像在视觉上更加平滑，丧失了真实世界中的偶然和不规则性。例如在上述讨论的扩散模型中，计算机通过于潜空间中执行去噪过程逐渐还原图像，这一过程中的算法利用了大量的概率推断和统计学原理。在这一过程中，程序倾向于选择更“常见”或“平均”的特征来生成图像，从而最大限度减少噪音与不确定性。然而，这种处理也导致了图像细节的平凡化，使其缺乏真实世界中可能涌现的独特性和不规则特征。

总而言之，AIGI不仅未消除CGI在技术物耦合阶段图像的诸种问题，更将这些问题以更加致密的技术黑箱封装于人类难以知觉的区域。如果说电子游戏图像以一种拒绝思考“在手性”的姿态规训人类对CGI的认识，那么人工智能时代的计算机生成图像则以一种主动开放与接纳的姿态向人类发出邀约：现象学还原的断裂并不可怕，对技术乌托邦的承诺可以超越一切幻觉。至此，CGI从欺骗的图像走入了谄媚的图像。

### 三、平滑世界谎言：从欺骗到谄媚

“平滑是某种单纯惹人喜爱的东西。它没有对立的否定性，不再是对抗体。如今，交际也为了使人们能够顺利地交换信息而变得平滑。平滑的交际中没有任何对他者与异者的否定。同者之间的相互回应，会使交流达到最高速度。”<sup>[24]</sup>

韩炳哲在其著作《美的救赎》中展现了一个“平滑世界”的基本样貌。这个世界充满了透明性和无摩擦的表面，消解了矛盾和差异，使人们沉浸于无止境的绝对积极的知觉体验中。显然，平滑世界是规训社会与消费社会结合的必然结果，且这种平滑指向着多重意味。

首先是现象学层面的平滑。界面成为“围绕文化和媒介的话语的中心舞台”<sup>[25]</sup>，光滑的界面即一切，我们只需操心界面提供的便利而无需考虑界面之下凛冽的真实——这里的“无需”同样包含着两种否定性经验，第一种是被上文以电子游戏图像为例讨论的CGI所建立的象征秩序直接否定的，换言之一种需要强制执行的否定性经验；第二种则是以AIGI为例讨论的现象学还原断裂诱发的，一种将“无”排除的状态，换言之一种无处也无物可否定的经验。

其次是意识形态层面的平滑。一如齐泽克在《图绘意识形态》一书中所评述的：

今天的社会必然是后意识形态的，而其中盛行的意识形态即犬儒主义；人们不再相信意识形态的真理，不再严肃的对待意识形态主张。然而，意识形态的基本层面不是遮蔽事物真实状态，而是构建社会现实本身的（无意识的）幻象。<sup>[26]</sup>

正如第一部分总结计算机生成图像的视知觉欺骗时所言，其许诺了一个“模拟世界”的乌托邦，但这个乌托邦包含着一种回溯性的否定——即对否定性的绝对抛弃，一种以玩家全情投入且毫无反思的沉浸作为前提才能诞生的乌托邦，这个平滑世界依靠一种对于差异的结构性的忽视——比如指云为草，比如永不可能发生的碰撞。韩炳哲对此也评述道：

数字化之美排除了他者的否定性。因此，它非常平滑不会有裂隙。其标志是不加任何否定的满足，即我喜欢。数字化之美营造了一个相同者的平滑空间，这里不可以存在陌异性与相异性。

而人工智能生成图像更进一步地强调这种否定性并非“被排除”，反而于生成的过程之中就“不存在”，甚至以一种积极姿态被纳入到我们的知觉体验——一种“本自具足”<sup>[27]</sup>的平滑。这是比强调对裂隙排除更为凛冽的状态，挪用一個数学术语来说，它将实在界的褶皱“卷积”了。如此说来，平滑世界并非某种寓言而是人类正在生存的现实。这样的演绎似乎导向某种晦暗的未来，但问题显而易见，平滑的究竟是世界吗？这跟我们观察者所在的位置有着直接关系。这一点我们可引入巴迪欧的事件哲学加以辨析：

我所谓的事件就是当某个完整的秩序（或规则）中留下某些无法被完全解释或处理的东西，这些东西（剩余）就是事件。事件不是给定的，因为所有能够给予或确定的事物，都是这个完整秩序的一部分。因此，事件实际上是我们通过解读和理解所产生的结果。<sup>[28]</sup>

纳入到我们所讨论的CGI中，便是在其致密且平滑中，总会涌现无法被客观把握的真实（随机、噪音、错

误……)，被这种真实在本体论上所撕裂的裂缝就是事件。有趣的地方就在于此，事件让不在这一层面上的情势被迫溢出：

在我们对诸多之多进行计数时，不可避免的至少会产生一个不属于原来的多或集合的元素，这个元素就是“溢出”，它对其他的多进行计数，唯独不对自己进行计数。对于这个多我们无法在原来的属于的层次上将其把握……只能在  $n+1$  上才是可能的……溢出点表明绝对连贯的不可能性。<sup>[29]</sup>

正如学者蓝江所述，理论上面对平滑世界的诸种溢出只能在更高层次上把握，但由于事件的不可避免，它们以意外的姿态被识别了。这种情形在不同的理论中都有相似的表达，诸如“创伤性时刻”、“实在界对象征界的入侵”、“谵妄”、“失神”等等。但这里要注意的是，相比于其他情形，巴迪欧重新整理了事件的发生学机制：“从本体论的架构上来说，所有的情势都是历史性情势，历史性的诸多无所不在”<sup>[30]</sup>。换句话说，巴迪欧建立的事件哲学从元结构出发证明了事件存在的必然性：一种“非结构的结构性存在”<sup>[31]</sup>。一种内蕴于结构的积极断裂，无法在既有结构中被整合继而溢出，带来新的可能与新的秩序。

正是在这一符号秩序下，事件的溢出显得尤为重要。尽管 CGI 试图通过其高度平滑的生成机制构建一个连贯、致密的感知世界，但正如巴迪欧的事件哲学所指出的，这种平滑的世界总会被不可避免的“剩余”所撕裂。CGI 及其背后的技术机制，在力求呈现一个无缝连接的虚拟现实时，恰恰结构性地制造了种种无法被控制的噪音与错乱，暴露出符号体系的局限。这里，事件的涌现不仅表明 CGI 作为“世界模拟器”无法达到真实的全面性，还揭示出人类对这一技术幻象的依赖，这种依赖很大程度来自图像自身从技术物到数码物不断建立的符号秩序。有趣的是，这种秩序首先从知觉层面出发给予人类种种限制与否定性规训，又在人工智能的参与下将种种限制转变为谄媚——肯定性积极。其所承诺的世界模拟其实指向对“人类中心主义的主动谄媚”。

需要指出，这与许多经典的有关图像谄媚的理解不同。不论是 W.J.T 米切尔 (W.J.T Mitchell) 评述威廉·霍加斯 (William Hogarth) 关于如画 (Picturesque) 的分析时所表达“吸引观者的主动参与”<sup>[32]</sup>，还是布雷德坎普的图像行为理论，他们强调的更多是一种可被看作图像行为 (Bildakt) 的图像的能动性，即图像通过它的存在、形式和构造，对世界施加某种行为。但在 CGI 这里，谄媚的图像是一种“外翻的图像”<sup>[33]</sup>，一种颠倒了持存生成机制的图像，这种肯定性积极是自我递归的，是无法、不能也无所谓与世界发生关系的。正是在这点上，CGI 图像的谄媚达到了一个新的层次：它的自我递归和积极性肯定使得它不再寻求与外部世界的互动，而是通过对符号的不断自我复制，形成了一种积极的空转。在这种谄媚中，CGI 根本不在模拟世界，而是通过其高度技术化的生成过程，试图让观者全然地笼罩在符号秩序构建的幻象中。

总之，作为一种技术神话，CGI 体现着人类从规训社会走向功绩社会的文化症候。人类贪婪的投向高速流通的 CGI，在一场人机共谋的狂欢中编织更加致密的符号秩序，并披上名为“世界模拟器”的虚假外衣。

### 结语：倦怠图像的例外抗争

齐泽克曾在《欢迎来到实在界这个大荒漠》中写道：“阿兰·巴迪欧认为，20 世纪的一个主要特征便是‘对实在界充满激情’”<sup>[34]</sup>。我们可以挪用这句话——21 世纪的一个主要特征便是“对图像充满激情”。正如韩炳哲在《倦怠社会》中论述的那样，激情的另一面即是倦怠。谄媚的图像也是倦怠的图像。被技术调校得完美无缺的图像在世间无尽流转，反映着人类个体的自我规训与耗散，在循环与自指的无尽链条中吟唱着“模拟世界”的虚假歌谣。当然，有趣的事情再次发生，对模拟世界的诘问突然在“去人类中心主义”的地方展开。2017 年一篇主题为“一

像素攻击(One Pixel Attack)”的论文,提供了一种只靠修改数字图像中的某一个像素,便可让人工智能将其错误的识别为另一形象的具体实践方法。<sup>[35]</sup>这一实践意外地揭示了谄媚图像所固有的缺陷,这些难以被人类知觉的缺陷在“人工智能的同类”那里仿佛一碰就碎的泡沫。在这无人参与的意外时刻,倦怠的图像却为人类找到了抵抗的可能。🚩

(责任编辑:陈曦)

#### 注释:

- [1] Sora 技术报告 [EB/OL]<https://openai.com/index/video-generation-models-as-world-simulators/>.
- [2] 缸中之脑假设将一个大脑从人体取出,放入装有营养液的缸里维持生理活性,超级计算机向大脑传递神经电信号并给予完全反馈,则大脑所体验世界其实是计算机的模拟,故此大脑能否意识到自己生活在虚拟现实之中?
- [3] Putnam H. *Reason, Truth and History*[M]. Cambridge: Cambridge University Press, 1981 :1-21.
- [4] Hartmann S. *The World as a Process: Simulations in the Natural and Social Sciences*[M]. Modelling and simulation in the social sciences from the philosophy of science point of view. Dordrecht: Springer, 1996: 77-100.
- [5] Bostrom N. Are we living in a computer simulation? [J]. *The Philosophical Quarterly*, 2003, 53(211): 243-255.
- [6] Brecht B. *The radio as an apparatus of communication*[M]. New Media. Routledge, 2003: 29-31.
- [7] (德国)弗里德里希·基特勒著.李双志译.实体之夜[M].上海:上海社会科学院出版社, 2019 : 27.
- [8] Umirzokov R R, Mambetov R M. ORIGIN AND DEVELOPMENT OF COMPUTER GRAPHICS[J]. *Мировая наука*, 2019 [10 (31)]: 17-18.
- [9] (德国)克劳斯·皮亚斯著.熊硕译.电子游戏世界[M].上海:复旦大学出版社, 2021-11 :89.
- [10] Bogost, Ian. *Persuasive games: the expressive power of video-games* [M]. Cambridge: MIT Press, 2007. 4.
- [11] 这里结合了拉康的“实在界”与德勒兹的“褶皱”概念,可以被理解为不可完全被符号化中复杂、连续缠绕与堆叠的结构。
- [12] 具体技术细节可参考 <http://www.coranac.com/tonc/text/mode7.html>.
- [13] 精灵图:当一张二维图像集成进场景中,成为整个显示图像的一部分时,这张图就称为精灵图。
- [14] 耿游子民.模组的铭刻与分离:电子游戏影像的诞生[J].文化艺术研究, 2023, (04): 44-53+114.
- [15] 李洋.作为工具性体验的电子游戏——及其与电影的关系[J].文化艺术研究, 2023, (04): 36-43+113.
- [16] Cohen H. The further exploits of AARON, painter[J]. *Stanford Humanities Review*, 1995, 4(2): 141-158.
- [17] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. *Advances in neural information processing systems*, 2020, 33: 6840-6851.
- [18] 位图:一种由像素组成的图像,每个像素都包含了颜色和位置信息。
- [19] CLIP 使用大量“图像—文字对”加以训练用以将文本与图像关联,在此处作为文本特征引导的噪音介入去噪过程。
- [20] 现象学还原是现象学中的一个核心概念,指通过去除所有关于外部世界的预设和先入之见回归到对主观经验本身的直接考察。它强调的是通过主观意识来把握经验世界的本质。
- [21] (法国)吉尔伯特·西蒙东著.许煜译.论技术物的存在模式[M].南京:南京大学出版社, 2024.
- [22] 耿游子民,陈旭光.从符号的自生成到去同步观看——CGI 技术下的影游融合新美学[J].未来传播, 2020, 27 (04): 54-60.
- [23] DeepDream 是由谷歌工程师 Alexander Mordvintsev 设计的计算机视觉程序,它使用卷积神经网络通过算法幻想性错觉来增强图像。
- [24] (德国)韩炳哲著.关玉红译.美的救赎[M].北京:中信出版集团股份有限公司, 2019 : 11.
- [25] Galloway A R. *The interface effect*[M]. Cambridge: Polity, 2012:38-39.
- [26] Zizek S. *Mapping Ideology*[M]. London : Verso, 1994 : 316.
- [27] 佛教用语,此处强调一种先验性丰满。
- [28] Badiou, Alain. *Can Politics Be Thought?* [M]. Durham: Duke University Press, 1985 : 78.
- [29] 蓝江.忠实于事件本身:巴迪欧哲学思想导论[M].北京:北京师范大学出版社, 2018 : 167-168.
- [30] (法)阿兰·巴迪欧.存在与事件[M].南京:南京大学出版社, 2024 : 233.
- [31] 可以这样理解,事件虽然无法结构化,但事件位是一种结构性存在。类比计算机里的数据与指针,数据虽然可以没有结构,但指针是一种典型的结构。
- [32] 这段论述来自 2024 年 4 月 20 日 W.J.T 米切尔于博雅国际电子游戏年会上所做题为《游戏之境:地点、视点与情境》的主题演讲。
- [33] 肖熹,李洋.什么是游戏影像[J].电影艺术, 2021, (05): 18-24.
- [34] (斯)齐泽克著,季广茂译.欢迎来到实在界这个大荒漠[M].南京:译林出版社, 2014 : 1.
- [35] Su J, Vargas D V, Sakurai K. One pixel attack for fooling deep neural networks[J]. *IEEE Transactions on Evolutionary Computation*, 2019, 23(5): 828-841.