

未来视界

在人类认知与文明进程中,图像始终扮演着重要角色。从史前洞穴壁画到现代数字影像,图像表达与生成方式的演进,深刻反映了人类认识世界与改造世界的深度与广度。

施展了一场持续三十年的「视觉魔法」
上海市数字媒体处理与传输重点实验室
在「视界」展现「看见」的实力



■ 国内首套高清电视地面广播样机系统

本报记者 易蓉

1907年,英语中首次出现“television”一词,用来描述一种通过无线电波将远方景象实时传送到观看者眼前的理论系统。“tele-”意味着“远”,“vis”是“看”,二者结合,便定格“电视”的技术构想。

在中国工程院院士、上海交通大学讲席教授张文军领衔的团队,三十年来的工作正是围绕图像(vision)的“处理”与突破距离(tele-)的“传输”展开,使“把远处景象传到眼前”这件事变得更清晰、更稳定、更自由,并最终在“视界”的竞争里展现中国“看见”的实力。

从“雪花屏”到“上海芯”

对于许多70后和80后而言,儿时电视屏幕上的“雪花”和需要手动调整的“天线”是共同记忆。那是模拟信号的时代,画面不稳定、色彩寡淡、分辨率低,且难以进一步提升。

“进入数字化时代,一切有了新的可能。”张文军说。实验室承担的首个国家级重大攻关项目便是研制我国第一套高清电视数字编码传输系统。这套系统就像给电视信号修了一条高速、保真的“数字公路”,将节目编码、传输到接收的全过程数字化,让接收清晰度比普通电视高出4倍,画面无闪烁,音响如激光唱盘般纯净。

该系统在1999年国庆50周年庆典上完成实况转播——这是中国数字电视从“无”到“有”的历史性一刻。而2000年国庆,市民在上海国际会议中心首次目睹数字高清电视转播的国庆节目,惊叹“灵呀!”。连连赞叹背后,是中国电视信号从模拟到数字的历史性跨越。

让每个家庭都看到高清电视,实现真正产业化必须攻克核心芯片。“芯片从无到有,从小到大,在十几年间经历了无数次艰难的技术攻关,但现在,我们已实现了自主化。”实验室里,张文军将一枚只有二分之一指甲盖大的芯片放在记者手心里时感慨地说。

2004年,实验室自主研发的三块“上海芯”问世,涵盖了节目接收解调、信号解码、图像格式转换三大数字电视接收最关键环节。这些芯片和实验室ADTB-T系统单载波技术,让上海市民第一次在飞驰的车上也能看到“不抖不卡”的高清画面,连时速430公里的磁悬浮列车上,信号接收成功率都接近100%。这些重大突破确保了我国数字电视标准的性能超越了国外同类标准。

从“跟着跑”到“定规矩”

2010年起,团队敏锐抓住电视与互联网融合的契机,首创了电视广播与互联网视频协同传输的系统架构,让广播电视的“广覆盖、高可靠”与互联网的“互动性、个性化”融合。这一架构支撑了5G广播、超高清直播、应急广播等国家重大应用,神舟飞船返回舱画面能实时高清传回、偏远山区在灾害中仍能收到应急信息的背后都有这套系统的影子。

中国是全球电视机出口大国,国际数字电视市场长期被欧美标准把持,形成极高的专利壁垒。为了扭转我国电视工业在专利领域受制于人的被动局面,必须掌握国际标准中最有价值的、最核心、不可绕开的“标准必要专利”。

话题主持:本报记者 易蓉

在这场演进的长河中,人类在19世纪发明了摄影术,在100年前发明了电视(Television),二者都深刻影响着文明的传播。而今天,AI正重新定义图像——图像不仅是对现实的投射,也是对世界

的思想衍射。

图像即世界。

这个世界里,像素颗粒是如何存在、如何流淌、如何构建成可见的图景的?今天,就让我们来拉开“视界”的科技帷幕。

图像未来

上海交通大学图像通信与网络工程所 副所长 宋利

图像的数字化,是这场演进长河中的“大事件”。

从最微观的视角来解构数字化图像,科学家发明了“像素”。在学术领域,与“摄取”像素相关的技术称之为“图像处理”,与“创制”像素相关的技术归为“计算机图形学”。也有两者交叉而成新的方向,比如“计算摄像学”。总结来说,计算摄像学是“主动调控硬件去拍,让后续处理更简单”;图像处理是“对已拍的图做加工,让图像更好看/更有用”;计算机图形学是“从无到有造图像,模拟真实或虚拟世界”。

人工智能技术作为通用型的基座技术,无论对“图像处理”还是“计算机图形学”以及“计算摄像学”,都产生了深刻的影响和技术革新甚至是颠覆。特别是2025年,图像AI生成技术(包括文生图、图生图、图生视频等)取得飞速进步,Nano Banana、GPT-Image、Sora、Kling、Wan等AI生图工具的陆续推出并迭代版本。AI进行图像创作,已经涌现非常多的爆火作品,甚至进入平台生产力工具池,帮助普通公众进行创意和创作。

AI生图的第一波热潮起始于2014年的生成式对抗网络(GAN),但存在训练不稳定,输出质量差等问题,并未转化为产品级可实际应用的工具。到2020年首次系统化提出的去噪扩散概率模型(DDPM,简称扩散模型),具有理论基础扎实、训练稳定性好、高质量数据生成等优点,成为当前热潮的底层关键技术驱动力。随着算法和工程持续优化,以及大语言模型(LLM)的加持,该技术已经日趋成熟,正全面走向商业落地应用。

而在规模法则(Scaling law)的驱动下,AI生图技术还在持续进步,正在从单帧图片到连续视频,再到可交互的三维空间,并保持风格、逻辑和物理规律的统一。图像模型正从纯文本的自由创意生成转向对象级甚至像素级的特定任务,即从“多样性”到“可控性”;生图工具不只看懂文字,更能通过多种模态输入、人机多轮对话等方式理解创作者的意图,即从“下命令”到“共创作”;多模态大模型正将“文字—图像—视频—3D”无缝串联,并致力于用统一架构同时处理多种模态的“理解—生成—剪辑—推论”任务,即从“单点强”到“全域优”。

目前AI领域的热词,如空间智能、物理智能、世界模型等都直接或间接地包含了图像(视频)生成这一关键“步骤”。因此,AI生图也是通往AGI(通用人工智能)的关键垫脚石。目前,上海交通大学图像通信与网络工程所正在开展生成式媒体通信、多模态大模型压缩、全真数智人及具身智联体等前沿技术攻关与创新,通过将AI与图像技术、通信技术的深度联姻,探索出具备“真像摄取+仿真计算”“真实场景+虚拟生成”“高效重建+实时驱动”特征的世界模型。

围绕“看见”,显示终端也在图像数字化浪潮中不断更新迭代,如今VR头盔、AR眼镜、MR等新型显示终端也日趋常见,将图像直接投影至网膜,甚至更进一步通过脑机接口在视觉皮层上“投射”图像的技术也已问世。完全可以期待在不远的未来,更多奇妙设备将来到大家眼前。

作为重要未来方向之一,扩展现实(XR)将带来无限想象力和新鲜体验。它以任意形式混合叠加现实环境和虚拟环境,使来自物理世界“摄取”的具象表达与来自思想世界“创制”的想象表达,经由像素颗粒自然汇集,融合一体。XR内涵丰富,是不断拓展图像技术边界的“理想国”。自由视角观看体育比赛、博物馆和演艺场景中的真3D体验(包括点云、体视以及全息等多种形态)已经进入应用。而神经辐射场(NeRF)、高斯激波(GS)等新型XR表征技术,支持通过手机、消费类的AI眼镜等设备进行实时图像的采集以及日常训练素材的积累;在AI的加持下,3D转换技术正在快速落地,已经涌现出Polycam、Remy等手机App,能将日常照片转换为高质量的3D模型,特别适合捕捉精细物体和场景。

未来3—5年,随着内容的丰富和终端的普及,XR将带动空间智能相关的各种生活应用(如室内导航,景区导览,数字化身等)走进寻常百姓家。