中国科学: 信息科学

SCIENTIA SINICA Informationis

论文

AIGC:NLP 在内容生成领域的全新发展

田晶怡 [1]

1. 中国海洋大学, 青岛 266400
Email:tianjingyi55@gmail.com

摘要 近年来，ChatGPT 以及 Sora 等大模型的出现，让人工智能领域的研究热点逐渐聚焦于 AICG。作为解决人类与计算机之间语言交流问题的关键技术，NLP 的研究对于 AICG 是必不可少的。尽管在 AIGC 这个充满希望和快速发展的研究领域中已经提出了各种各样的方法，但据我所知，鲜少有相关人员努力系统地总结这些工作以及阐述其与 NLP 之间的紧密关联。为了给未来工作的发展奠定基础，在本文中，我试图通过对最近 AIGC 方法进行广泛的回顾来填补这一空白。具体来说，1)我首先提出现有的 AIGC 大模型以及相对应的生成技术;2)然后，我概述了与 AIGC 相关的库，包括常用数据集;3)最后，我讨论了 AIGC 与 NLP 技术的紧密联系和目前研究面临的一些关键挑战，并分享了我对未来 NLP 技术在 AIGC 领域应用潜在方向的见解。

关键词 人工智能生成，自然语言处理，应用，算法与模型，发展

1 介绍

AIGC（Artificial Intelligence Generated Content）即由人工智能生成的内容，涵盖了文本、图像、音频到视频等多种形式的创作，均由 AI 系统独立生成。AIGC 的出现和发展，为生产出大量、高质量的内容提供了全新的解决方案，同时也在创意领域和内容产业中掀起了一股变革。NLP（Natural Language Processing，自然语言处理）技术在 AIGC 中扮演着至关重要的角色。它作为一项人工智能技术，负责解决人类与计算机之间的语言交流问题。在 AIGC 中，NLP 技术被广泛应用于生成自然语言文本，辅助 AI 系统理解、分析和生成各种形式的文字内容。具体地说，NLP 技术可以通过文本生成模型、文本摘要算法、实体识别、情感分析等方法，帮助 AI 系统地

生成高质量的文本内容。通过 NLP 技术，AI 可以理解语言的语法、语义和语境，从而生成与人类写作相媲美甚至超越的文本作品。

在 AIGC 的技术架构中，三个核心要素不可或缺：数据、硬件及算法。其中，音频、文本与图像等高质量信息素材构成了算法训练的基础模块，其规模与来源直接关联到预测结果的精确度。在硬件层面，尤其是计算资源的配备，构成了 AIGC 技术实施的基础设施支撑。鉴于对计算力的需求日益增加，高速、高效的芯片技术以及云计算方案的融合变得尤为关键，它们共同应对于处理海量参数与数据流的挑战。在此框架下，算法的效能直接决定了内容产出的品质，而优质的数据资源与高性能硬件的协同支持，对于优化模型输出结果发挥着决定性作用。因此，探究这三者之间的动态互动与优化路径，对于推动人工智能生成技术的学术研究和应用发展具有重要意义。

接下来讨论当前热门的在线图像生成平台。作为最先进的在线图像生成技术之一，Midjourney 已发展至第 5.2 代，为用户提供了付费使用的便捷途径。与此同时，DALL·E，这一由 OpenAI 公司精心研发的平台，同样在业界赢得了极高的声誉。用户支付费用后，可以将 DALL·E 3 模型与 ChatGPT Plus 版、企业版无缝结合，实现更为丰富的图像生成体验。除此之外，微软公司还将 DALL·E 模型融合进了 Bing 的聊天功能中，使得用户在浏览器中就能享受到该前沿技术的便利。同时，Stability AI 公司研发的 Stable Diffusion 平台已经发布了包括 v2.0、v2.1 在内的多个开源预训练模型版本。用户不但可以下载开源模型在本地进行计算，也可以通过在线接口使用，充分满足了不同用户的需求。

上述产品都是强大的 text-to-image 工具，用户输入的文本内容，它们便能够快速生成与之相匹配的高质量图像。此外，Pixeling、wukong 等在线平台还特别支持中文输入的 text-to-image 功能，进一步拓宽了用户的使用范围。

最近几年 AIGC 的发展突飞猛进，可访问性也达到了前所未有的高度。尽管不少国内国外的研究学者提出了很多高质量的 AIGC 大模型，但最近的只有少数工作试图全面总结评价这些成果，仍然缺乏对这一新兴邻域的进展和面临挑战的系统回顾。为了填补空白，本文总结了 AIGC 基于的模型以及其关键技术（第 2 节）,概述库和介绍常用训练集(第 3 节)和讨论未来的研究方向(第 4 节)。本文的目的是为了让对该领域具有兴趣的研究人员能够对 AIGC 有较为全面的了解并获得该领域的最新进展。

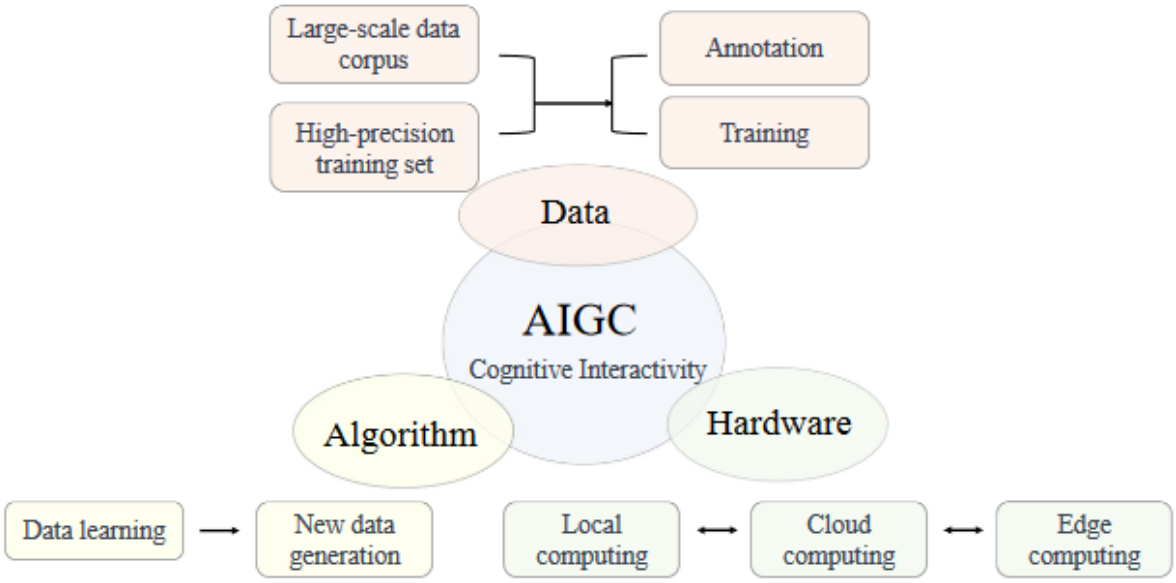

图 1 AIGC 中硬件、算法和数据之间的关系

2 模型阐述

在 AIGC 中，生成式模型和语言模型常常结合使用，以实现更复杂和多样化的内容生成任务。例如，结合生成对抗网络和语言模型可以生成更加逼真和多样的文本内容；结合生成式图像模型和语言模型可以实现图像描述生成等应用。这些结合使用的模型能够更好地模仿人类创作的过程，生成更具创造性和多样性的内容。

生成式模型是指一类机器学习模型，可以根据输入数据生成新的数据。这些模型可以学习数据的分布并生成符合该分布的新数据。在 AIGC 中，生成式模型通常用于生成各种类型的内容，如文本、图像、音频等。常见的生成式模型包括生成对抗网络（GAN）、变分自动编码器（VAE）等。

语言模型是一种用于衡量句子或文本序列出现概率的模型。它可以根据给定的前文预测接下来可能出现的文本内容。语言模型在 AIGC 中通常被用来生成文本内容，如自动写作、对话生成等。常见的语言模型包括循环神经网络（RNN）、长短时记忆网络（LSTM）、Transformer 等。

2.1 生成式模型

在 AIGC 这个概念尚未如此名声大噪之前，图片生成的应用已经存在，比如利用 GAN 网络进行 AI 换脸等。而近几年 AI 绘图和 ChatGPT 等大规模语言模型（LLMs）分别在两个领域表现出惊人的效果并广受关注，AIGC 这一概念才开始被大家熟知。目前使用较为广泛的图片生成模型主要有 3 种基础架构，分别是对抗生成网络 GAN 系列（Generation Adversarial Network）、变分自动编码器 VAE 系列（Variational Automatic Encoder）和扩散模型 DM 系列

（Diffusion Model）。其中 AI 绘图以 2020 年的去噪扩散概率模型 DDPM（Denoising Diffusion Probabilistic Model）[11]为一个较大的里程碑，在此之前的生成模型主要以 GAN 居多。目前使用最多、效果最好的开源 AI 绘画模型 Stable Diffusion 则为扩散模型，据悉 MidJourney 是变形注意力 GAN 的变体。

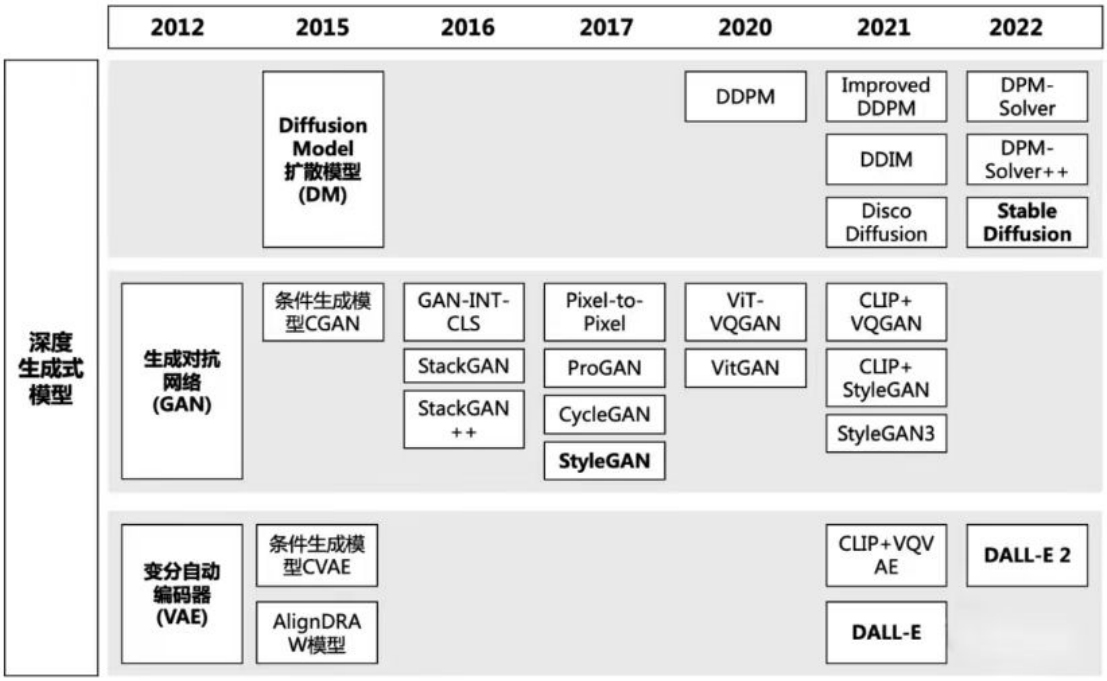

图 2 生成式模型总览

2.1.1 GAN:生成对抗网络简介

GAN(Generative Adversarial Network)即生成对抗网络，其实质可理解为两个互补网络的组合。具体而言，生成器（Generator）负责模拟数据的创建，而鉴别器（Discriminator）则负责鉴别输入数据的真伪。生成器的目标是使其生成的数据足够逼真，以至于鉴别器难以辨别；而鉴别器则致力于提升自身的鉴别能力，以求更为精准地判断数据的真实性。

以上二者均是基于神经网络的架构。该模型的核心在于通过对网络内部神经元间权重参数的调整来实现优化。在这个过程中，生成器和鉴别器利用各自的损失函数，结合误差反向传播（Backpropagation）算法以及随机梯度下降法、牛顿形法等优化方法，不断地进行参数的调整和优化，进而提升整个 GAN 的性能。

生成网络的损失函数：

$$\boldsymbol{L_G = H(1, D(G(z)))} \quad (1)$$

上式中，G 代表生成网络，D 代表判别网络，H 代表交叉熵，z 是输入随机数据。D(G(z)) 是对生成数据的判断概率，1 代表数据绝对真实，0 代表数据绝对虚假。H(1,D(G(z)))代表判断结果与 1 的距离。

判别网络的损失函数：

$$\boldsymbol{L_D = H\big(1, D(x)\big) + H\big(0, D\big(G(z)\big)\big)} \quad \textbf{(2)}$$

上式中，x 是真实数据，这里要注意的是，H(1,D(x))代表真实数据与 1 的距离，H(0,D(G(z)))代表生成数据与 0 的距离。

根据公式推断，识别网络如果要想取得良好的效果，就要满足真实数据与 1 的距离小，生成数据与 0 的距离小。

2.1.2 VAE:变分自编码简介

VAE 模型，全称为变分自编码器，是一种基于神经网络的生成模型。它通过训练得到两个关键函数：推断网络与生成网络。这两个网络协同工作，能够生成在原始输入数据中未曾出现的新数据。VAE 模型在生成各种复杂数据类型方面表现出强大的能力，包括手写数字、人脸、门牌号、CIFAR 图像、场景中的物体、分割图像以及基于静态图像的预测等。接下来描述 VAE 的训练过程：

VAE 的损失函数:

$$\boldsymbol{minLoss_{VAE} = D_{KL}\big(q(z|x)\big|\big|P(z)\big) - E_{q(z|x)}[logP(x|z)]} \quad (3)$$

上式中，x 是输入数据,z 是隐变量，q(z|x)是给定输入 x 后 z 的后验分布，P(z)是 z 的先验分布，$\boldsymbol{D_{KL}\big(q(z|x)\big|\big|P(z)\big)}$是 q(z|x)和 P(z)之间的 KL 散度，用来衡量两个概率分布间的差异；$\boldsymbol{E_{q(z|x)}[logP(x|z)]}$是关于后验分布 q(z|x)的期望值，logP(x|z)是 z 输入 x 时的对数似然，这个期望值表示了生成模型 P(x|z)的重建能力，即模型能够准确地根据潜在表示 z 重建输入数据 x 的能力。

2.1.3 Diffusion Models：生成扩散模型简介

Diffusion Models（扩散模型）是通过逐步向训练数据中引入高斯噪声，进而学习数据恢复的过程，训练完成后，这些模型可将随机噪声样本输入，通过学习的去噪过程来生成新数据，本质上，扩散模型属于隐变量模型范畴，它借助马尔可夫链（Markov Chain, MC）将数据映射至隐空间[11]。在每一个时间步 t 中逐渐将噪声添加到数据 x_i中以获得后验概率 $q(x_{1:T}|x_0)$，其中$x_1, \cdots, x_T$ 代表输入的数据同时也是 隐空间。也就是说 Diffusion Models 的 隐空间与输入数据具有相同维度[11]。Diffusion Models 分为正向的扩散过程和反向的逆扩散过程,以下分别介绍：

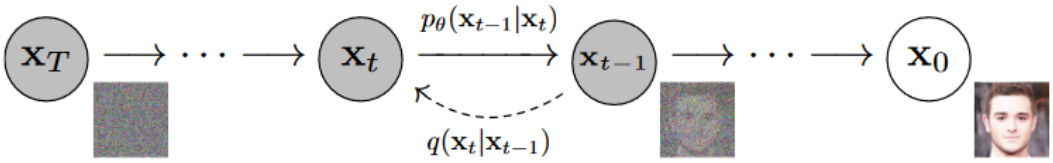

图 3 DM 的扩散和逆扩散过程

扩散过程：

$$q(x_{1:T}|x_0) \coloneqq \Pi_{t=1}^{T} q(x_t|x_{t-1}) \coloneqq \Pi_{t=1}^{T} N\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \quad (4)$$

上式中，$\beta_1,\ldots,\beta_T$是高斯分布方差的超参数，随着 t 的增大，x_t越来越接近纯噪声。当 T 足够大的时候，x_T 可以收敛为标准高斯噪声 $N(0, I)$。

逆扩散过程：

$$p_\theta(x_{T:0}) \coloneqq p(x_T)\Pi_{t=1}^{T} p_\theta(x_{t-1}|x_t) \coloneqq p(x_T)\Pi_{t=1}^{T} N(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t)) \quad (5)$$

根据马尔可夫规则表示，逆扩散过程当前时间步 t 只取决于上一个时间步 t-1，所以有：

$$p_\theta(x_{t-1}|x_t) \coloneqq N(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t)) \quad (6)$$

通过以上公式，我们已对 Diffusion Models 的扩散与逆扩散过程有了初步理解。然而，在马尔科夫具体实现的过程中，求解步骤往往复杂且关键。研究员们通常采用蒙特卡洛方法来进行采样，随后对得到的结果的有效性进行评估。这一步骤对于确保模型的准确性和效率至关重要。

2.1.4 flow 模型

NICE 模型是学习一个非线性双射转换（bijective transformation），其将训练数据映射到另一个空间，在该空间上分布是可以因子化的，整个模型架构依靠直接最大化 log-likelihood 来完成。

利用 change of variable 方法。对于转换 h=f(x)，我们假定 f 是可逆的，h 的维度和 x 的一样。训练目标函数如下[12]:

$$log(px(x)) = \sum_{d=1}^{D} log\left(pH_d(f_d(x))\right) + log\left(\left|det\left(\frac{\partial f(x)}{\partial x}\right)\right|\right) \quad (7)$$

2.2 大语言模型

Transformer 以其独特的自注意力机制，革新了传统深度学习模型，但仅限于文本序列建模。然而 ViT 模型的诞生打破了这一限制，成功地将 Transformer 架构应用于计算机视觉（CV）领域，实现了文本与视觉间的桥梁搭建。在此基础之上，BEiT 模型则进一步拓展了研究视野，将生成式预训练技术引入 CV 领域，为该领域的研究开启了新篇章。CLIP 的多模态模型的问世，则实现了图文两种模态间的交互，为图文互生性研究提供了新的可能性。此外，扩散模型与多模态大模型的结合，为文本生成图像领域的发展注入了新活力。

2.2.1 Transformer

Transforme 由 Google 在 2017 年提出基于自注意力机制，用于处理从序列到序列的建模任务。在长距离依赖和并行技术方面具有显著优势，是目前在自然语言处理领域中最先进的建模技术。

Transformer 由编码器和解码器构成，两者均由多层结构组成，每层又包含多头自注意力机制和全连接前馈网络。特别地，

其中的交叉注意力机制，在跨模态建模中发挥着关键作用，特别是在文图扩散模型中，得到了广泛应用。[1]

2.2.2 ViT

ViT 是 2020 年 Google 团队提出的将 Transformer 应用在图像分类的模型，但是因为其模型简单高效，可扩展性强，成为了 transformer 在 CV 领域应用的里程碑著作，打通了 Transformer 与 CV 领域的壁垒，从此图文可以统一建模。

ViT 将输入图片分为多个 patch（16x16），再将每个 patch 投影为固定长度的向量送入 Transformer，后续 encoder 的操作和原始 Transformer 中完全相同[1]。但是因为对图片分类，因此在输入序列中加入一个特殊的 token，该 token 对应的输出即为最后的类别预测[1]。

2.2.3 GPT 系列

Generative Pre-trained Transformer（GPT）系列是由 OpenAI 推出的 GPT 系列模型，作为一类卓越的预训练语言模型，其在复杂的 NLP 任务中展现了出色的性能。GPT 系列模型的核心价值在于，通过增强模型容量和语料规模，实现了模型能力的显著提升。

GPT-1：无监督预训练和有监督微调

GPT-1 采用了无监督的预训练方式，基于 transformer 架构，通过自回归机制预测后续词汇，以此捕捉语言的内在结构。随后，该模型在有标签的数据集上进行有监督的微调，以适应具体的 NLP 任务。这种无监督预训练与有监督微调的策略，对 NLP 乃至计算机视觉（cv）领域均产生了深远影响。

GPT-2：多任务学习模型

GPT-2 旨在解决 GPT-1 在有监督学习方面的局限性，通过自监督预训练实现多任务学习，无需依赖大量有标签数据。GPT-2 在模型结构上并未做大幅改动，而是通过增加网络参数和扩大无标签数据集规模，利用前缀编码技术实现多任务处理。这一改进使 GPT-2 在多项任务上的性能超越了传统的有监督学习方法。

GPT-3：少样本学习模型

GPT-3 则进一步引入了 In Context Learning 机制，实现了通过少量示例或指令快速学习新任务的能力。该模型能够依据自然语言指令或任务示例进行预测，完成新任务的学习。GPT-3 以其庞大的模型参数（最多达 1750 亿）和卓越的性能，推动了大规模预训练技术的发展，显著提高了模型在少样本、零样本场景下的表现，为通用人工智能的发展奠定了坚实基础。

2.2.4 T5 模型

T5，即 Transfer Text-to-Text Transformer，是谷歌于 2019 年推出的一个大型语言学模型，其参数规模约达 110 亿。在 AIGC 领域中，T5 常用于文本编码，并以其卓越的全局语义理解能力，在性能上超越了 CLIP 文本编码器。“Transfer”在此指代的是迁移学习（Transfer Learning），而 Text-to-Text 则构成了一个将各种 NLP 任务转化为从文本到文本的迁移学习形式的统一的框架。举例来说，翻译、情感分类、相似度计算和文本摘要等任务，均可以通过添加相应的任务前缀，利用 T5 模型进行预测。

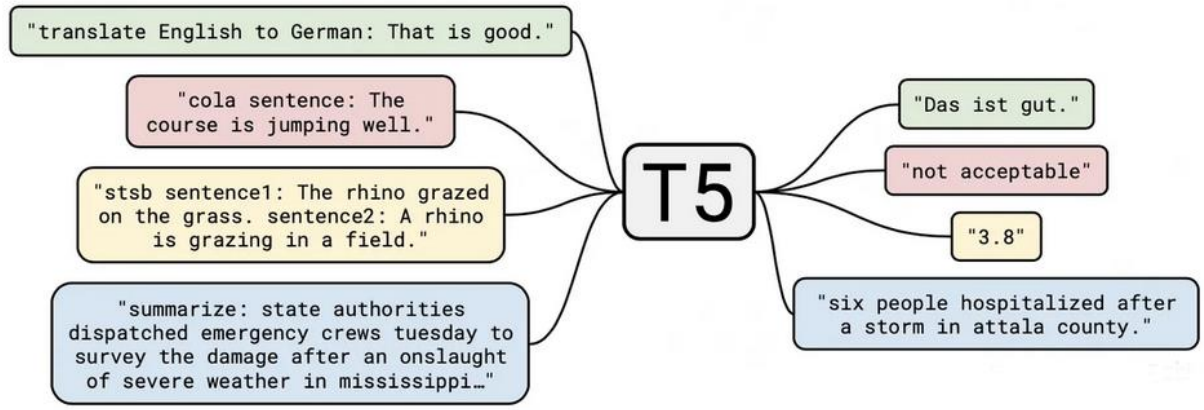

图 4 T5 模型的结构

T5 模型的学习策略包含以下几个方面：

a. 采用了 Transformer 的编码器-解码器（Encoder-Decoder）架构；

b. 借鉴了 BERT 的破坏（Masking）方法；

c. 实施了 Replace Span 的破坏策略；

d. 在数据破坏比例上设定为 15%；

e. 破坏时的小段长度被设定为 3。

以上学习策略共同作用下，T5 模型在多项任务上均取得了显著的性能提升，达到了该领域的顶尖水平（State-of-the-Art, SOTA）。此外，T5 模型还进一步验证了大型模型的重要性，即使参数规模达到 110 亿，模型的性能提升仍未出现减缓迹象，预示着更大的潜力等待发掘[17]。

3 相关数据集

鉴于当前图像生成和文本数据集已被广泛下载应用，本文重点探讨统计 3D 生成数据集。AIGC-3D 依据算法的不同特点可大致被分为三类：首先是基于 2D Stable Diffusion Model 的 Zero-shot 生成，其次是单图像驱动的 3D 生成，最后是拓展至 3D 数据的 Stable Diffusion Model 生成。

本文系统地整理了以下几类数据集：首先是 **Zero-Shot SD 3D** 生成数据集，它在训练和推理过程中均无需额外 3D 模型作为训练数据，而是直接利用 SD 预训练模型进行分数蒸馏（score distillation）；其次是**单图像 3D** 生成数据集，它在训练时需要输入单幅图像作为参考，其本质更接近于基于单幅图像的三维重建任务；再次是**基于 3D 数据的 Diffusion** 生成数据集；此外，还包括**多视角**数据集以及其他类型的 3D 数据集。这些数据集为 AIGC-3D 领域的研究提供了宝贵的资源。

表 1 AIGC-3D 最新公开数据集整理汇总

数据集名称	简介	适用类	地址
DreamFusion	基于 image parameterization 思想，率先提出 SDS loss	ZS 3D	https://dreamfusion3d.github.io/
Magic3D(Nvidia)	多分辨率 SDS	ZS 3D	https://www.nvidia.cn/omniverse/synthetic-data/
Latent-Nerf	潜在空间的 SDS	ZS 3D	https://github.com/eladrich/latent-nerf
Fantasia3D	几何和纹理耦合	ZS 3D	https://fantasia3d.github.io/
ProlificDreamer	提出 VSD loss,比 SDS 能够生成更多具有多样性和高质量的 3D 内容	ZS 3D	https://ml.cs.tsinghua.edu.cn/prolificdreamer/
SweetDreamer	用于多模态机器学习，特别是视觉、文本和音频任务	ZS 3D	https://sweetdreamer3d.github.io/
DreamGaussian	结合 SIGGRAPH23 的 3d-gaussian-splatting 做模型重建，模型重建效率提升	ZS 3D	https://dreamgaussian.github.io/
Make-It-3D	通过使用 2D 扩散模型作为 3D-aware 先验，从单个图像中创建高保真度的 3D 物体	单图像 3D	https://make-it-3d.github.io/
Ditto-Nerf	基于扩散模型的迭代式文本到三维图形全方位生成	单图像 3D	https://janeyeon.github.io/ditto-nerf/
ShapeNet	包含 3D 模型的组成结构，有模型组件的语义分割，3D 数据保存为 obj+mtl 格式，能够支持的纹理精细程度有限，有简单文本标注	基于 3D 数据的 DM	https://shapenet.org/
Objaverse	800k 静态 3D 模型，44k 带动画的 3D 模型，比 ShapeNet 有更精细的纹理，3D 数据保存为 glb 格式，可以有较复杂的纹理和光照信息，同时包含模型的 caption 文本标注	基于 3D 数据的 DM	https://objaverse.allenai.org/

UniG3D	结合 Objaverse 和 shapenet，对这两个数据集中的数据做了 normalization 和 alignment	基于 3D 数据的 DM	https://unig3d.github.io/
OminiObject3D	无 caption 文本标注	基于 3D 数据的 DM	https://omniobject3d.github.io/
ScanNeRF	NeRF 重建 benchmark，数据组成为多视角采集图像+GT NeRF 模型，无 caption 文本标注	多视角数据	https://eyecan-ai.github.io/scannerf/
Co3D	大规模的、多样化的、基于视频的 3D 目标检测数据集	多视角数据	Reizenstein_Common_Objects_in_3D_Large-Scale_Learning_and_Evaluation_of_Real-Life_ICCV_2021_paper.pdf
GSO	包含了一系列的场景图片，以及与之相对应的图形结构信息，这些图形结构信息反映了场景中物体的几何关系和拓扑结构。	多视角数据	https://arxiv.org/pdf/2204.11918.pdf
NAVI	该数据集由 36 个对象的多视图和野生图像集合组成，总共约有 10K 个图像。	多视角数据	https://navidataset.github.io/
ModelNet40 (Princeton)	CAD volumetric 数据	其他	http://3dvision.princeton.edu/projects/2014/3DShapeNets/
3D Future	家具数据集	其他	https://tianchi.aliyun.com/specials/promotion/alibaba-3d-future

4 讨论

4.1 联系

在探讨 AIGC 与 NLP 技术的关联性时，我们不能忽视 NLP 技术对于文本数据的深度解析、精细处理及创新生成起的关键作用，正是依赖于这些 NLP 技术，AIGC 才能能够产出高度逼真的文本内容。AIGC 的核心应用领域广泛，包括但不限于文本创作、摘要提取、机器翻译和情感分析等多个方面，而正是凭借对 NLP 技术的精确运用，AIGC 系统方能准确解析并生成既符合语言规则又具备深层语义的文本内容。

以新闻报道、学术论文、日常对话及叙事性故事的生成为例，NLP 技术不仅确保了生成文本的语法准确性，更在语义连贯性、主题贴合度以及风格一致性上达到了极高的水准。近年来，AIGC 的迅猛发展得益于深度学习驱动的生成式 AI 技术的迭代更新以及海量的（多模态）训练数据的支持。

俞士纶教授团队在最新的综述论文[13]中，对自 2014 年以来的单模态与多模态（特别是图像-文本多模态）领域中的生成式 AI 技术进行了详尽的梳理。从 NLP 的视角看，生成式 AI 技术经历了从基于 N-Gram 的神经语言模型，到循环神经网络（LSTM、GRU 等）的广泛应用，再到 Transformer 架构下的各种先进模型（如 ELMo、GPT、BART、T5 等）的崛起，其模型复杂度与性能均超越了计算机视觉领域。

而在计算机视觉领域，生成式 AI 技术则从 GAN、VAE、Flow-based 等传统图像生成技术起步，逐渐演进至 StyleGAN、VQVAE 等更为先进的模型，极大地提升了图像生成的品质。近年来，随着扩散模型与 Vision Transformer（ViT）技术的崛起，为图像生成领域注入了新的活力。

值得注意的是，多模态图像文本生成式 AI 技术作为近年来的研究热点，通过融合图像与文本两种模态的数据，能够生成更加真实、丰富的图像或文本内容，这也预示着未来 AIGC 发展的一个重要方向。

4.2 机遇

AIGC 技术的广泛应用为各行各业注入了新的活力。在科研领域，图扩散模型已渗透至分子结构、蛋白质分析和材料科学等多个分支，推动了研究的深入；在医疗健康领域，该技术不仅支持生成和共享保护隐私的合成临床健康数据，还助力聋哑人群实现语音生成，促进了医疗服务的个性化与人性化；在艺术领域，基于 GAN 的多尺度生成模型和创新的损失函数为自动修复和上色提供了可能，同时，数据增强技术也通过生成更多训练数据，为文献[17]等提供了有效的替代和补充。相较于传统的人工临摹上色，这种 AI 辅助的修复方式显著提升了效率；在商业与办公领域，OpenAI 推出的 ChatGPT 模型以及微软逐步推出的 365 Copilot 和 Windows Copilot 等智能工具，能够高效处理大量低智重复的工作，并为复杂任务提供有力支持，极大地提升了办公自动化的水平，进而提高了整体的工作效率。

4.3 挑战

研究方面。我们重点关注 AIGC 模型在文本生成过程中的多样性与真实性。为了避免生成重复且不连贯的内容，我们致力于优化模型性能，尤其在长文本生成任务中，更要留意信息流失和语义不一致的问题。此外，NLP 领域的数据往往伴随着噪声和标签不精确的问题，如何妥善解决这些问题，对于 AIGC 模型的训练与生成效果具有举足轻重的意义。

应用方面。生成模型在训练过程中涉及对用户数据的学习，引发了大众对隐私保护的关切。有研究表明，生成数据中可能存在的隐私泄露风险可通过某些对抗性攻击被揭露。因此，我们必须对此保持高度警惕。在安全层面，AIGC 模型生成内容的真实性与其内容的潜在危害性和实用性难以完全掌控，这使得 AIGC 的规范化之路充满挑战。为了确保生成内容的安全与合规，我们需要不懈努力。再者，版权问题亦不容忽视。当前，AIGC 的版权归属在法律层面尚不清晰，国际间对此亦未有统一定论。这需要我们进一步探讨与明确，以确保相关权益得到有效保障。

4.4 未来方向

多模态。AIGC 的未来发展中，多模态融合是一个显著的趋势。这种融合不仅限于文本和图像等数据，还将有助于拓展 AIGC 在自然语言处理领域的应用范围，从而为用户带来更全面和深入的信息体验。当前，文本图像生成技术已相对成熟，而未来，我们有望融合更多模态数据，构建出功能更为强大的 AIGC 模型。

文本生成模型的革新。预计在未来的发展中，通过引入更为复杂的 AIGC 架构，结合注意力机制和迁移学习等先进技术，我们可以提升文本生成模型的性能，使其生成的内容更具多样性和真实性。

强化学习与生成模型的结合。 为了提升 NLP 任务的生成效果，我们可以尝试将强化学习技术融入生成模型的训练中。这种结合可以指导生成模型更精准地满足特定需求，从而获得更为出色的生成性能。

专业化发展。 尽管 AIGC 的基础模型是通过在大量通用数据上进行预训练来获取知识的，但在面向如医疗健康等专业化应用时，仍需收集一定量的专用数据进行任务型的微调。如何在确保专业化的同时，减少对专用数据的依赖，是 AIGC 未来发展的一个重要方向。

标准化和规范化。 为了确保 AIGC 技术的健康发展，我们需要完善其输出控制的标准化机制，并加强与 AIGC 相关的法律法规建设，如用户隐私保护、产品版权等。通过这些措施，我们可以促进 AIGC 与人类社会的和谐共生，并防止技术的滥用，最终实现负责任的 AIGC 技术应用。[21]

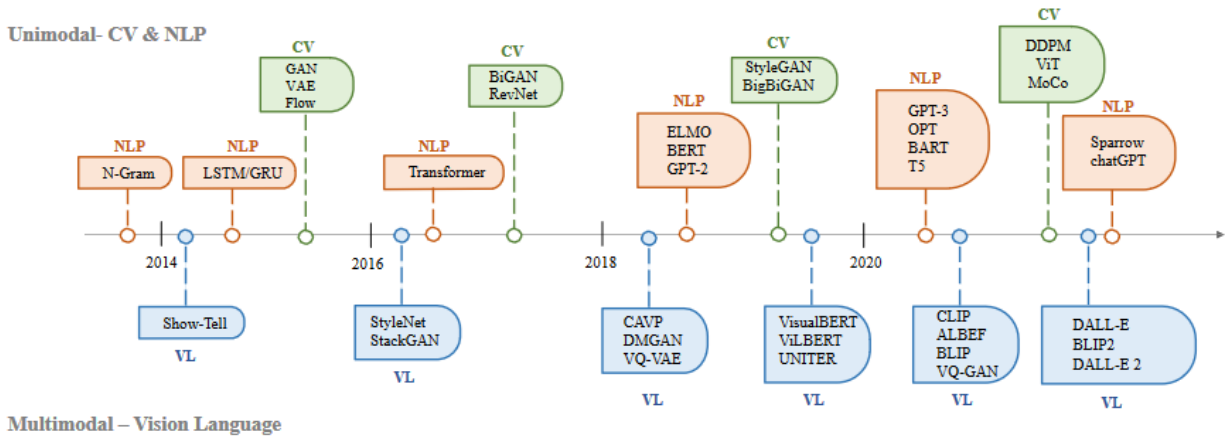

图 5 生成式 AI 的发展史

参考文献

1 Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, et al. A Survey of Large Language Models. arXiv:2303.18223, 2023

2 Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, et al. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258, 2021

3 Chaoning Zhang, Chenshuang Zhang, Sheng Zheng, Yu Qiao, et al. A Complete Survey on Generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 All You Need? arXiv:2303.11717, 2023

4 Gerhard Paaß, Sven Giesselbach. Foundation Models for Natural Language Processing -- Pre-trained Language Models Integrating Media. arXiv:2302.08575

5 Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, et al. Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. arXiv:2304.01852, 2023

6 Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, Hong Lin. AI-Generated Content (AIGC): A Survey. arXiv:2304.06632, 2023

7 Chaoning Zhang, Chenshuang Zhang, Chenghao Li, Yu Qiao, Sheng Zheng, Sumit Kumar Dam, et al. One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era. arXiv:2304.06488, 2023

8 Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, Lichao Sun. A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. arXiv:2303.04226, 2023

9 Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, Xia Hu. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. arXiv:2304.13712, 2023

10 Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, et al. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. arXiv:2302.09419, 2023

11 Jonathan Ho, Ajay Jain, Pieter Abbeel. Denoising Diffusion Probabilistic Models. arXiv:2006.11239, 2020

12 L Dinh, D Krueger, Y Bengio. NICE: Non-linear Independent Components Estimation. In ICML, 2014

13 https://en.wikipedia.org/wiki/Illiac_Suite

14 Y. Cao et al., “A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT.” arXiv, Mar. 07, 2023. doi: 10.48550/arXiv.2303.04226.

15 M. Zhang et al., “A Survey on Graph Diffusion Models: Generative AI in Science for Molecule, Protein and Material,” 2023.

16 B. Zhou, G. Yang, Z. Shi, and S. Ma, “Natural language processing for smart healthcare,” IEEE Rev. Biomed. Eng., pp. 1–17, 2022, doi: 10.1109/RBME.2022.3210270.

17 腾讯研究院，AIGC 发展趋势报告 https://cloud.tencent.com/developer/article/2255694

18https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

19 G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein, “Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models.” arXiv, Dec. 12, 2022. doi: 10.48550/arXiv.2212.03860.

20 C. Zhang et al., “One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era.” arXiv, Apr. 04, 2023. Accessed: May 27, 2023. [Online]. Available: http://arxiv.org/abs/2304.06488

21 https://towardsdatascience.com/a-pathway-towards-responsible-ai-generated-content-6c915e8155f9

22 Lago F, Pasquini C, Böhme R, et al. More real than real: A study on human visual perception of synthetic faces [applications corner][J]. IEEE Signal Processing Magazine, 2021, 39(1): 109-116.

23 Wang S Y, Wang O, Zhang R, et al. CNN-generated images are surprisingly easy to spot... for now[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8695-8704.

24 Frank J, Eisenhofer T, Schönherr L, et al. Leveraging frequency analysis for deep fake image recognition[C]//International conference on machine learning. PMLR, 2020: 3247-3258.

25 Liu Z, Qi X, Torr P H S. Global texture enhancement for fake face detection in the wild[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 8060-8069.

26 Ju Y, Jia S, Ke L, et al. Fusing global and local features for generalized ai-synthesized image detection[C]//2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 3465-3469.

27 Liu B, Yang F, Bi X, et al. Detecting generated images by real images[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 95-110.

28 Tan C, Zhao Y, Wei S, et al. Learning on Gradients: Generalized Artifacts Representation for GAN-Generated Images Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 12105-12114.

29 Ojha U, Li Y, Lee Y J. Towards universal fake image detectors that generalize across generative models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 24480-24489.

30 Wang Z, Bao J, Zhou W, et al. DIRE for Diffusion-Generated Image Detection[J]. arXiv preprint arXiv:2303.09295, 2023.

31 Zhong N, Xu Y, Qian Z, et al. Rich and Poor Texture Contrast: A Simple yet Effective Approach for AI-generated Image Detection[J]. arXiv preprint arXiv:2311.12397, 2023.

32 Zhu M, Chen H, Yan Q, et al. GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image[J]. arXiv preprint arXiv:2306.08571, 2023.