

# 澳洲幸运10到底有没有规律

EMCm7DuGMf9IBRLV

澳洲幸运10到底有没有规律UniToken：多模态“全能选手”，一次编码搞定图文理解与图像生成

首次在统一框架内实现理解与生成的“双优表现”，打破了多模态统一建模的僵局！

复旦大学和美团的研究者们提出了UniToken——一种创新的统一视觉编码方案，在一个框架内兼顾了图文理解与图像生成任务，并在多个权威评测中取得了领先的性能表现。

UniToken通过融合连续和离散视觉表征，有效缓解了以往方法中“任务干扰”和“表示割裂”的问题，为多模态统一建模提供了新的范式。

为了便于社区内研究者们复现与进一步开发，UniToken团队已将代码与模型全部开源。

在传统图文理解或图像生成模型中，其视觉编码的底层特性差异较大。

譬如图文理解模型（如LLaVA、Qwen-VL等）要求从图像中抽取高层语义，从而进一步结合文本进行协同理解；而图像生成模型（如DALL-E、Stable Diffusion等）则要求保留充分的底层细节以高保真图像的生成。

由此，开发理解生成一体化的多模态大模型面临着以下几大难题：

**视觉编码割裂**：理解任务偏好具有高层语义的连续视觉特征（如CLIP），而生成任务依赖保留底层细节的离散视觉特征（如VQ-GAN编码的codebook）；

**联合训练干扰**：理解与生成任务差异而带来的冲突性使得在统一模型中训练时难以兼顾两个任务的性能，存在“一个优化，另一个退化”的现象。

为了应对上述挑战，领域内的相关工作通常采取两类范式：以VILA-U等为代表的工作通过结合图像重建与图文对比学习的训练目标，来提升离散视觉编码的语义丰富度；以Janus等为代表的工作通过为理解和生成任务分别定制相应的视觉编码器与预测头，来实现两个任务之间的解耦。

然而，前者在理解任务上目前依旧难以与连续视觉编码驱动的多模态大模型匹敌；后者则在应对更复杂的多模任务（例如多轮图像编辑等）时面临严重的上下文切换开销及单边信息缺失等问题。

不同于Janus的多任务解耦的设计思路，UniToken为所有下游任务均提供一套完备的视觉信息，促使多模态大模型以指令驱动的形式从中吸收相应的知识。

具体而言，UniToken采取统一的双边视觉编码器，其中将VQ-GAN的离散编码与SigLIP的连续表征以下述方式进行拼接，从而得到一套兼备高层语义与底层细节的视觉编码：

[BOS][BOI]{离散图像token}[SEP]{连续图像embedding}[EOI]{文本}[EOS]

多阶段训练策略

为了协调理解与生成任务的特性，UniToken采用三阶段训练流程：

基于Chameleon作为基座，本阶段旨在为LLM接入SigLIP的连续视觉编码。为此，在训练时冻结LLM，仅训练SigLIP ViT和Adapter，使其输出与语言空间对齐。

基于第一阶段对齐后的双边编码器所提供的完备视觉信息，本阶段在大规模图文理解与图像生成数据集上联合训练，通过控制数据配比（10M:10M）以均衡提升模型理解与生成任务的性能。

通过测试发现，第二阶段训练后的模型在指令跟随、布局图像生成等方面的表现均有待加强，故在本阶段进一步引入高质量多模态对话（423K）与精细化图像生成数据（100K），进一步增强模型对复杂指令的跟随能力。

得益于保存了双边视觉编码的完备性，UniToken可无缝衔接现有的细粒度视觉增强技术。

具体而言，UniToken在连续视觉编码侧引入两项增强策略：

**AnyRes**：将高分辨率图像划分为多个子图，分别提取特征后进行相应空间位置的拼接，以提升对图像的细粒度感知；

**ViT端到端微调**：在模型的全训练流程中，动态微调连续视觉编码器的权重，结合精细的学习率控制策略以避免模型崩溃，进而适应广泛的场景。

**实验结果**：超越SOTA，多模态统一的“优等生”

在多个主流多模态基准（图文理解+图像生成）上，UniToken均取得了媲美甚至领先于领域内专用模型的性能：

与此同时，研究者们对于训练策略及视觉编码的影响进行了进一步深入的消融分析：

**总结**：迈向通用理解生成一体化的多模态大模型

从发展趋势上来看，目前图文理解模型在通用性上远远领先于图像生成模型。

而Gemini-2.0-Flash与GPT-4o在指令跟随的图像生成方面的惊艳表现，带来了通用图像生成模型未来的曙光。

在这样的时代背景下，UniToken仅是初步的尝试，而其信息完备的特性也为进一步挖掘其更深层次的潜力提供了更多信心：

**模型规模扩展**：借助更大的语言模型，进一步探索统一模型在理解与生成上的“涌现能力”；

**数据规模扩展**：引入更大规模的训练数据（如Janus-Pro使用的近2亿样本），推动模型性能极限；

**任务类型扩展**：从传统的理解与生成拓展至图像编辑、故事生成等图文交错的任务，追逐通用生成能力的上限。

论文链接：<https://arxiv.org/pdf/2504.04423> 代码地址：<https://github.com/SxJyJay/UniToken>

千里马计划(免费)

幸运168飞艇官网开奖网址app

澳洲10人工计划网

澳洲幸运5官网真实吗

168澳洲幸运10开奖官网网站

正规快三彩票app下载

正规大小单双快三下载软件

澳洲幸运5开奖官网开奖号查询

澳洲幸运10杀2码公式

澳州10开奖记录

澳洲幸运10精准计划老师qq

澳洲幸运10开奖号码查询

7码5期必中的倍投法

168澳洲幸运10开奖官网授权

全天飞艇免费计划最新版

澳洲幸运5官方开奖直播

168澳洲幸运5官网结果记录体彩

168澳洲幸运10开奖历史查询

2023澳洲幸运5开奖直播