

延伸阅读

模型介绍

可图大模型是由快手可图团队开发的基于潜在扩散的大规模文本到图像生成模型。Kolors 在数十亿图文对下进行训练，在视觉质量、复杂语义理解、文字生成（中英文字符）等方面，相比于开源/闭源模型，都展示出了巨大的优势。同时，Kolors 支持中英双语，在中文特色内容理解方面更具竞争力。更多的实验结果和细节请查看我们的[技术报告](#)。

评测表现

为了全面比较 Kolors 与其他模型的生成能力，我们构建了包含人工评估、机器评估的全面评测内容。在相关基准评测中，Kolors 具有非常有竞争力的表现，达到业界领先水平。我们构建了一个包含14种垂类，12个挑战项，总数量为一千多个 prompt 的文生图评估集 KolorsPrompts。在 KolorsPrompts 上，我们收集了 Kolors 与市面上常见的 SOTA 级别的开源/闭源系统的文生图结果，并进行了人工评测和机器评测。

人工评测

我们邀请了50个具有图像领域知识的专业评估人员对不同模型的生成结果进行对比评估，为生成图像打分，衡量维度为：画面质量、图文相关性、整体满意度三个方面。Kolors 在整体满意度方面处于最优水平，其中画面质量显著领先其他模型。

模型	整体满意度平均分	画面质量平均分	图文相关性平均分
Adobe-Firefly	3.03	3.46	3.84
Stable Diffusion 3	3.26	3.5	4.2
DALL-E 3	3.32	3.54	4.22
Midjourney-v5	3.32	3.68	4.02
Playground-v2.5	3.37	3.73	4.04
Midjourney-v6	3.58	3.92	4.18
Kolors	3.59	3.99	4.17

所有模型结果取自 2024.04 的产品版本

机器评测

我们采用 [MPS](#) (Multi-dimensional Human preference Score) 来评估上述模型。我们以 KolorsPrompts 作为基础评估数据集，计算多个模型的 MPS 指标。Kolors 实现了最高的MPS 指标，这与人工评估的指标一致。

模型	MPS综合得分
Adobe-Firefly	8.5
Stable Diffusion 3	8.9
DALL-E 3	9
Midjourney-v5	9.4
Playground-v2.5	9.8
Midjourney-v6	10.2
Kolors	10.3

更多的实验结果和细节请查看我们的技术报告。点击[技术报告](#)。

快速使用

要求

- python 3.8及以上版本
- pytorch 1.13.1及以上版本
- transformers 4.26.1及以上版本
- 建议使用CUDA 11.7及以上

1、仓库克隆及依赖安装

```
1 apt-get install git-lfs
2 git clone https://github.com/Kwai-Kolors/Kolors
3 cd Kolors
4 conda create --name kolors python=3.8
5 conda activate kolors
6 pip install -r requirements.txt
7 python3 setup.py install
```

2、模型权重下载 ([链接](#)) :

```
1 huggingface-cli download --resume-download Kwai-Kolors/Kolors --local-dir weights/Kolors
```

或者

```
1 git lfs clone https://huggingface.co/Kwai-Kolors/Kolors weights/Kolors
```

3、模型推理：

```
1 python3 scripts/sample.py "一张瓢虫的照片，微距，变焦，高质量，电影，拿着一个牌子，写着“可图”"# The image will be saved to "scripts/outputs/sample_text.jpg"
```

4、Web demo：

```
1 python3 scripts/sampleui.py
```

在 Diffusers 中使用

确保您安装了最新版本的 `diffusers` (0.30.0.dev0)：

```
1 git clone https://github.com/huggingface/diffusers
2 cd diffusers
3 python3 setup.py install
```

注意：

- `KolorsPipeline` 默认使用 `EulerDiscreteScheduler` 作为噪声调度器。我们推荐使用该调度器时搭配 `guidance scale=5.0` 及 `num_inference_steps=50`。
- `KolorsPipeline` 同时支持 `EDMDPMSolverMultistepScheduler`。在使用该噪声调度器时，推荐使用参数 `guidance scale=5.0` 及 `num_inference_steps=25`。
- 除了文生图能力，`KolorsImg2ImgPipeline` 同时也支持图文生图功能。

运行以下指令进行图像生成：

```
1 import torch from diffusers import KolorsPipeline pipe = KolorsPipeline.from_pretrained(
```

```
2     "Kwai-Kolors/Kolors-diffusers",
3     torch_dtype=torch.float16,
4     variant="fp16"
5 ).to("cuda")
6 prompt = '一张瓢虫的照片，微距，变焦，高质量，电影，拿着一个牌子，写着"可图"' image =
7 pipe(
8     prompt=prompt,
9     negative_prompt="",
10    guidance_scale=5.0,
11    num_inference_steps=50,
12    generator=torch.Generator(pipe.device).manual_seed(66),
13 ).images[0]
14 image.show()
```

协议、引用、致谢

协议

Kolors (可图) 权重对学术研究完全开放，如需商用请填写[问卷](#)，发送问卷至 kwai-kolors@kuaishou.com 进行登记获取许可后免费使用。

本开源模型旨在与开源社区共同推进文生图大模型技术的发展。本项目代码依照 Apache-2.0 协议开源，模型权重需要遵循本《模型许可协议》，我们恳请所有开发者和用户严格遵守[开源协议](#)，避免将开源模型、代码及其衍生物用于任何可能对国家和社会造成危害的用途，或用于任何未经安全评估和备案的服务。需要注意，尽管模型在训练中我们尽力确保数据的合规性、准确性和安全性，但由于视觉生成模型存在生成多样性和可组合性等特点，以及生成模型受概率随机性因素的影响，模型无法保证输出内容的准确性和安全性，且模型易被误导。本项目不对因使用开源模型和代码而导致的任何数据安全问题、舆情风险或因模型被误导、滥用、传播、不当利用而产生的风险和责任承担任何法律责任。

引用

如果你觉得我们的工作对你有帮助，欢迎引用！

```
1 @article{kolors,
2   title={Kolors: Effective Training of Diffusion Model for Photorealistic
3   Text-to-Image Synthesis},
4   author={Kolors Team},
5   journal={arXiv preprint},
```

```
5     year={2024}  
6 }
```

致谢

- 感谢 [Diffusers](#) 提供的codebase
- 感谢 [ChatGLM3](#) 提供的强大中文语言模型