

更多AI工具可直接访问：<https://www.faxianai.com/>

Deepseek满血版入口：<https://www.faxianai.com/ai/6039.html>

# 工具教程：Flux



作者：郑敏轩/Mike/六耳

敏轩

郑敏轩

[https://space.bilibili.com/3461579519560151?spm\\_id\\_from=333.999.0.0](https://space.bilibili.com/3461579519560151?spm_id_from=333.999.0.0)

## Flux制作团队由来：

Flux模型团队制作者：原stable diffusion核心团队因开源理念不合，出来成立“黑森林”团队。黑森林获3200万美元投资，拿出第一款产品：Flux

## 首次一次发布三种型号：

拿出来时有三种型号：

Flux pro（最好的，闭源使用）

Flux dev（开源，供学习研究使用，不支持商用）

Flux schnell（开源，开源最彻底，支持商用）

## 参数量比较：

Flux的参数训练量达到了12B，且语言支持自然语言。

12B我们类比一下8月6日的智谱发布的开源视频模型CogVideoX-2B，开源的是2B的一个模型，所以模型的视频推理结果不如快手的可灵。8月28日智谱最新开源了5B的视频模型。

视频还在5B，由此可见图像12B确实是一个高度（开源里面）

所以Flux一出现就占据极重要影响。



**Flux：训练参数12B（120亿）**

**SD3 Large：训练参数8B（80亿）**

**SD3 Medium：训练参数2B（20亿）**

## 硬件要求变高：

由参数带来的，是硬件要求变高，推理dev版本的Flux模型时，最低显存需要16G，建议24G，模型本身大小是22多G，如果训练的话，经过优化目前16G也可以训练Lora模型，但是DB方式训练微调模型的话，除显存最低要24G外，要求内存也需要32G以上才行。

## 模型特点：

图片质量审美进一步提升



手的崩坏得到优化



### 语言理解增强

由于训练时包含了文字，所以支持自然语言且理解“左右”等位置关系（早期SD1.5不支持不理解位置关系，所以几乎无法靠语言描述进行精确生成，现在对语言理解有一步不小的增强）且支持英文生成





仍不支持中文显示，但英文显示结果良好。

训练时就”见识“过各种尺寸大小：

所以生成图像的尺寸弹性很强，不像之前只能在512或1024、768得到好结果。

## 安装及使用：

因为上次**祐萌**刚刚讲过了comfyUI的使用，我们就讲一下在comfyUI中如何使用Flux

（在ForgeUI中同样可以使用，国内还可以去哩布哩布、阿里云等众多平台使用，有很多平台可以薅羊毛，这里介绍本地部署comfyUI）



把这 4 个模型，放到对应的目录下就可以了。

- t5xxl\_fp16.safetensors：放在 ComfyUI/models/clip/ 目录下
- clip\_l.safetensors：放在 ComfyUI/models/clip/ 目录下
- ae.safetensors：放在 ComfyUI/models/vae/ 目录下

- flux1-dev.safetensors：放在 ComfyUI/models/unet/ 目录下

## 怎么记和理解呢？

flux1-dev.safetensors是底模，ae.safetensors是VAE，clip\_l.safetensors和t5xxl\_fp16.safetensors是关于文字语言编码，这么理解就行。

## 哪里下载呢？

t5xxl\_fp16.safetensors 和 clip\_l.safetensors 下载地址：

[https://huggingface.co/comfyanonymous/flux\\_text\\_encoders/tree/main](https://huggingface.co/comfyanonymous/flux_text_encoders/tree/main)

t5xxl 分为 fp16 和 fp8，如果你内存超过 32GB，那就用 fp16 的，如果没超过，那就用 fp8 的。

ae.safetensors 和 flux1-dev.safetensors 下载地址：

<https://huggingface.co/black-forest-labs/FLUX.1-dev/tree/main>

我随后也准备一下百度网盘和夸克网盘。



更新：（下面准备了夸克和百度的网盘链接，方便部分同学下载）

flux相关模型（体积较大）的夸克网盘链接：<https://pan.quark.cn/s/b5e01255608b>

flux相关模型（体积较大）的百度网盘链接：

<https://pan.baidu.com/s/1mCucHrsfRo5SttW-03ei0g?pwd=ub9h> 提取码：ub9h

如果 GPU 性能不足、显存不够，底模可以使用 fp8 的量化版模型，速度会快很多，下载地址：

<https://huggingface.co/Kijai/flux-fp8/tree/main>

black-forest-labs **FLUX.1-dev** like 1.83k

Text-to-Image Diffusers Safetensors English FluxPipeline image-generation flux License: flux-1-dev-non-commercial-license (other)

Model card Files and versions Community

main FLUX.1-dev 6 contributors History: 7 commits + Contribute

|                               |                                   |           |                                               |             |
|-------------------------------|-----------------------------------|-----------|-----------------------------------------------|-------------|
| sumith1896 sayakpaul HF STAFF | Update README.md (#46)            | 3038751   | VERIFIED                                      | 4 days ago  |
| scheduler                     | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| text_encoder                  | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| text_encoder_2                | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| tokenizer                     | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| tokenizer_2                   | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| transformer                   | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| vae                           | Add diffusers format weights (#3) |           |                                               | 11 days ago |
| .gitattributes                |                                   | 1.66 kB   | Initial commit                                | 12 days ago |
| LICENSE.md                    |                                   | 16.9 kB   | Initial commit                                | 12 days ago |
| README.md                     |                                   | 4.27 kB   | Update README.md (#46)                        | 4 days ago  |
| ae.safetensors                |                                   | 335 MB    | Rename ae.sft to ae.safetensors               | 7 days ago  |
| dev_grid.jpg                  |                                   | 1.3 MB    | Initial commit                                | 12 days ago |
| flux1-dev.safetensors         |                                   | 23.8 GB   | Rename flux1-dev.sft to flux1-dev.safetensors | 7 days ago  |
| model_index.json              |                                   | 536 Bytes | Add diffusers format weights (#3)             | 11 days ago |

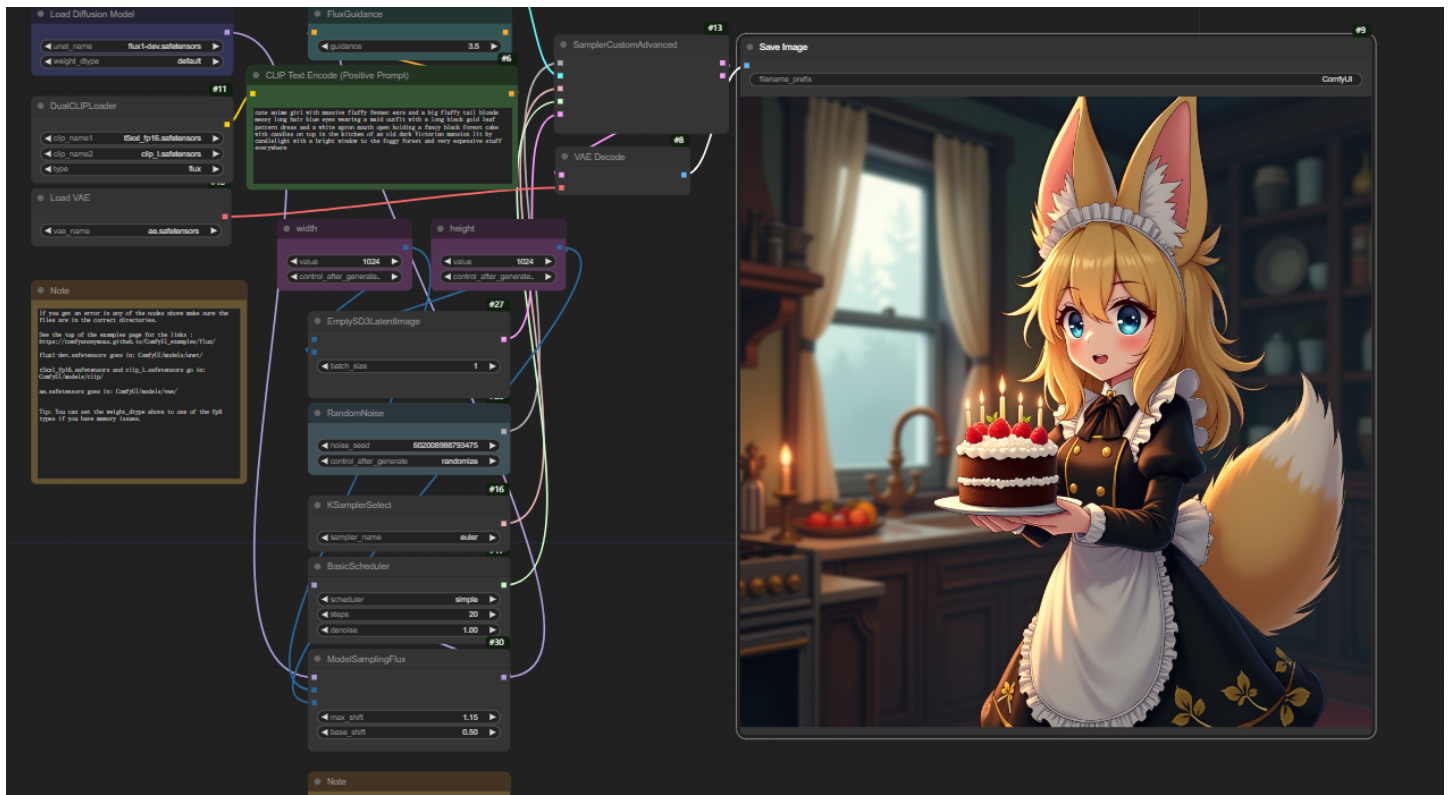
最后我们再下载dev的工作流：

 dev的官方原版workflow.json 14.97KB 

上面我把工作流复制上了，用上面这个就行。或者下面官方原版的图片链接，图片导入comfyUI就是工作流。

[https://comfyanonymous.github.io/ComfyUI\\_examples/flux/flux\\_dev\\_example.png](https://comfyanonymous.github.io/ComfyUI_examples/flux/flux_dev_example.png)

我们打开 ComfyUI，把工作流或图片拖拽到 ComfyUI 里：



## 显存不够16G？ 低配置方案？

开源社区迅速展开了对低配置方案的优化，NF4来自我们controlnet的作者，GGUF则包含多个版本可以使用

### NF4模型下载

<https://huggingface.co/lllyasviel/flux1-dev-bnb-nf4/blob/main/flux1-dev-bnb-nf4.safetensors>

放置在ComfyUI/models/checkpoint/中（不像其他 Flux 模型那样放置在UNET 中）

### NF4配套节点插件

git clone [https://github.com/comfyanonymous/ComfyUI\\_bitsandbytes\\_NF4.git](https://github.com/comfyanonymous/ComfyUI_bitsandbytes_NF4.git)

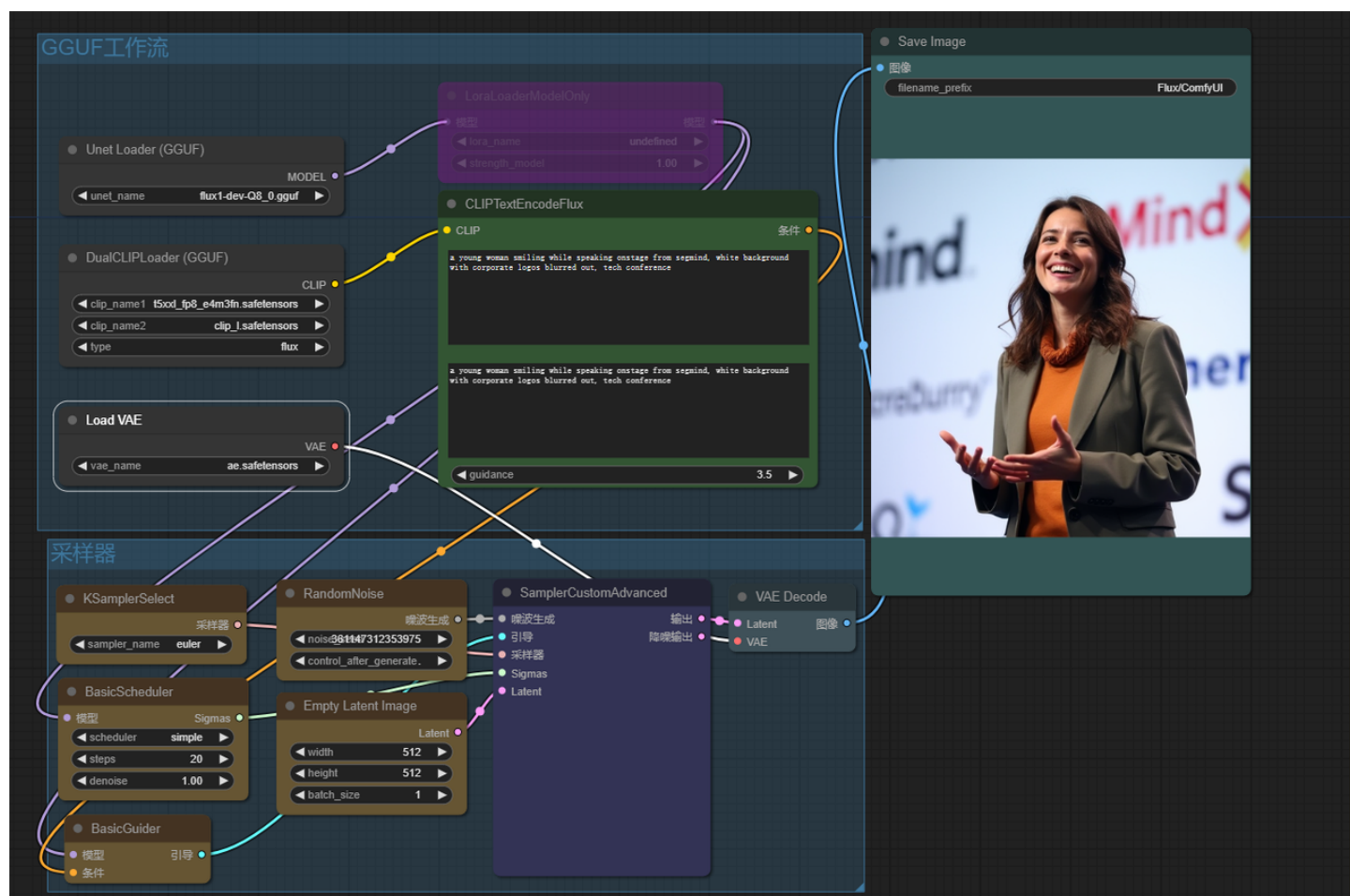
### GGUF模型下载

Flux GGUF模型：<https://huggingface.co/city96/FLUX.1-dev-gguf/tree/main>

### GGUF配套节点插件

GGUF节点包：<https://github.com/city96/ComfyUI-GGUF>

以下是使用GGUF生图：



值得一提的是在最新版本的comfyUI中GGUF的节点插件是可以在Manager管理器中搜到下载安装的，NF4的配套节点插件则搜不到。

Filter

All

GGUF

1 custom nodes

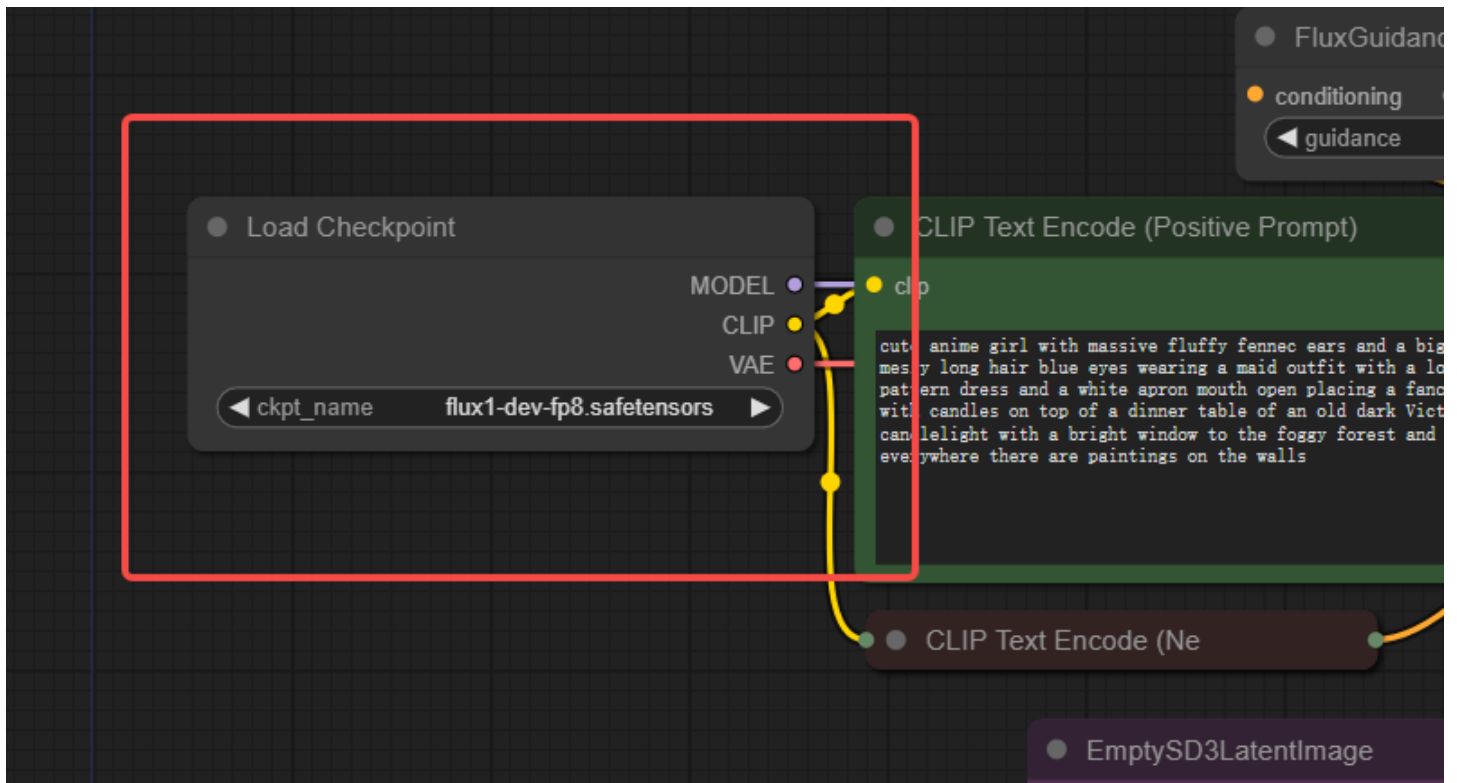
| <input type="checkbox"/> | ID | Title        | Install            | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Extensions                                                                          | Conflicts | Author                                                                                     | *   | Last Update |
|--------------------------|----|--------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------|-----|-------------|
| <input type="checkbox"/> | 54 | ComfyUI-GGUF | <div>Install</div> | <div><div>GGUF</div>Quantization support for native ComfyUI models. This is currently very much WIP. These custom nodes provide support for model files stored in the <div>GGUF</div> format popularized by llama.cpp.<br/><br/>While quantization wasn't feasible for regular UNET models (conv2d), transformer/DiT models such as flux seem less affected by quantization. This allows running it in much lower bits per weight variable bitrate quants on low-end GPUs.</div> |  |           |  city96 | 534 | 2024-08-26  |

注意使用精度优化的低配模型的话，工作流和原版是不一样的。此处没有专门列举。

**workflow.json**  
7.94KB

自己改的话就是把上面官方的这个fp8的工作流，只需把底模的节点换成NF4的或者GUFF的即可。





相关生态发展很快，有Lora、Controlnet、IP-adapter相关生态建设非常速度，以及字节最近发布的Flux Hyper lora是为了8步快速生图。

下节我们先讲讲Flux的lora训练。