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

郑敏轩： Flux与MiniCPM、Joy Caption的强强联合

作者 郑敏轩

Flux很强大，推高了开源图像生成质量，与此同时，优秀的MiniCPM和Joy Caption与Flux组合提供了强大工具的基础。

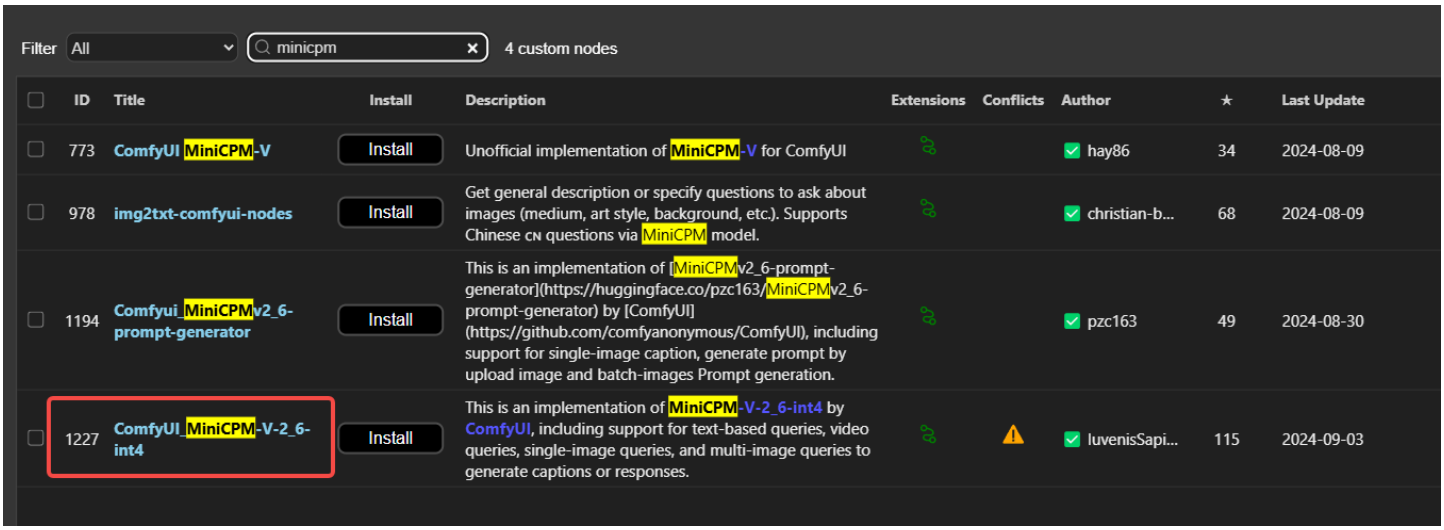
有测评说效果上florence2＜MiniCPM-V＜Joy Caption，部分表现Joy Caption会稍不如意，但整体Joy Caption是强的，这个过程中florence2稍显弱了，但是推理速度快。所以我们本章来说说MiniCPM-V和Joy Caption这两个。

在2024-09-19至2024-09-21的杭州云栖大会上，我们也见到了MiniCPM的身影：

MiniCPM-V

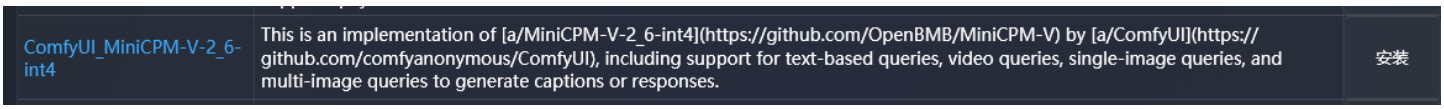
1. 话不多说，我们上节点：MiniCPM-V-2.6-int4

方法一：最简单的可以从节点管理器中安装（注意结尾是-V-2.6-int4的那个）：



在manager管理器中安装

方法二：或在秋叶包中安装（注意结尾是-V-2.6-int4的那个）：



或在秋叶包中安装

方法三：或者直接下载下面文件解压，复制ComfyUI_MiniCPM-V-2_6-int4文件夹到你的“\ComfyUI\custom_nodes”目录下。注意ComfyUI_MiniCPM-V-2_6-int4文件夹里面直接就是如下图的多个文件不能再包文件夹了。

夸克网盘：链接：https://pan.quark.cn/s/bc35e6c7e8a6

百度网盘：链接：https://pan.baidu.com/s/1sq9e2dcZsLGMDNNpmuYp6Q?pwd=jdei

提取码：jdei

2. 节点安装好后，还需要模型，

模型下载地址（解压后大小5.55G，压缩包大小4.85G）：

✍️ 夸克网盘：我用夸克网盘分享了MiniCPM-V-2_6-int4.rar

链接：https://pan.quark.cn/s/98c953d1ec8b

百度网盘：通过百度网盘分享的文件：MiniCPM-V-2_6-int4.rar

链接：https://pan.baidu.com/s/1y4wYyLn511al4LDEkIGEsA?pwd=bred

提取码：bred

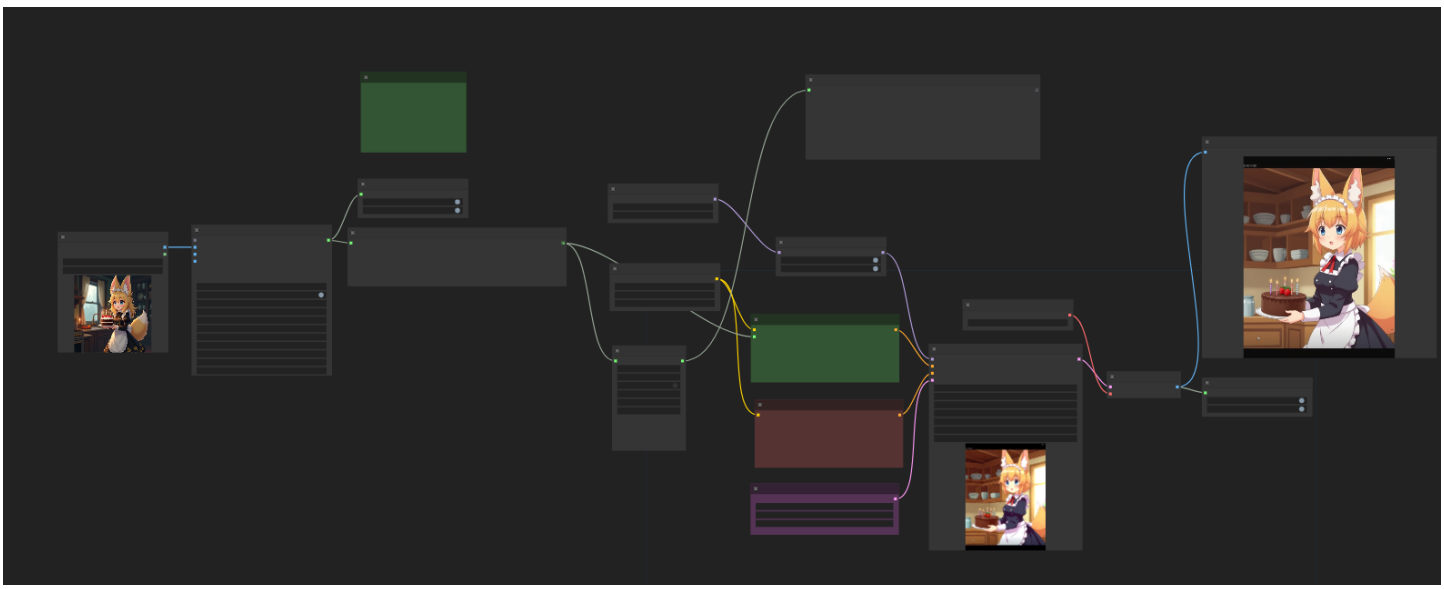
3. 模型下载好后放在：comfyui/models/prompt_generator/MiniCPM-V-2_6-int4
4. 导入基础工作流验证

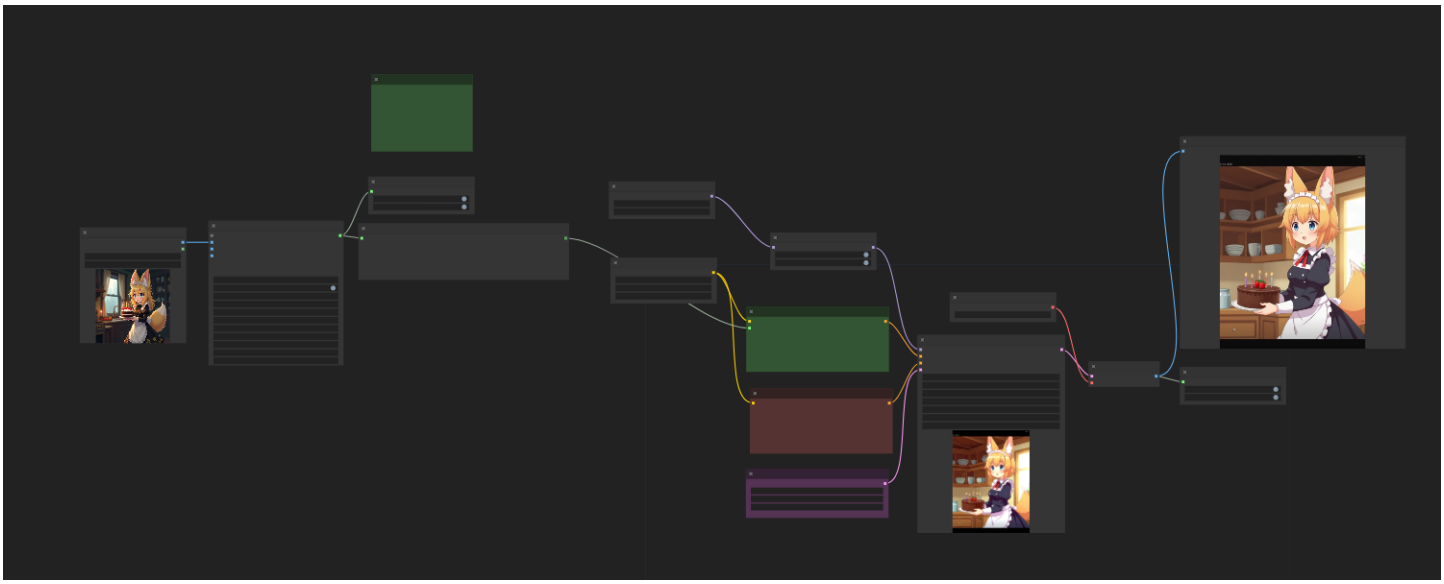
Chat_with_single_image_workflow_polished.json
4.77KB

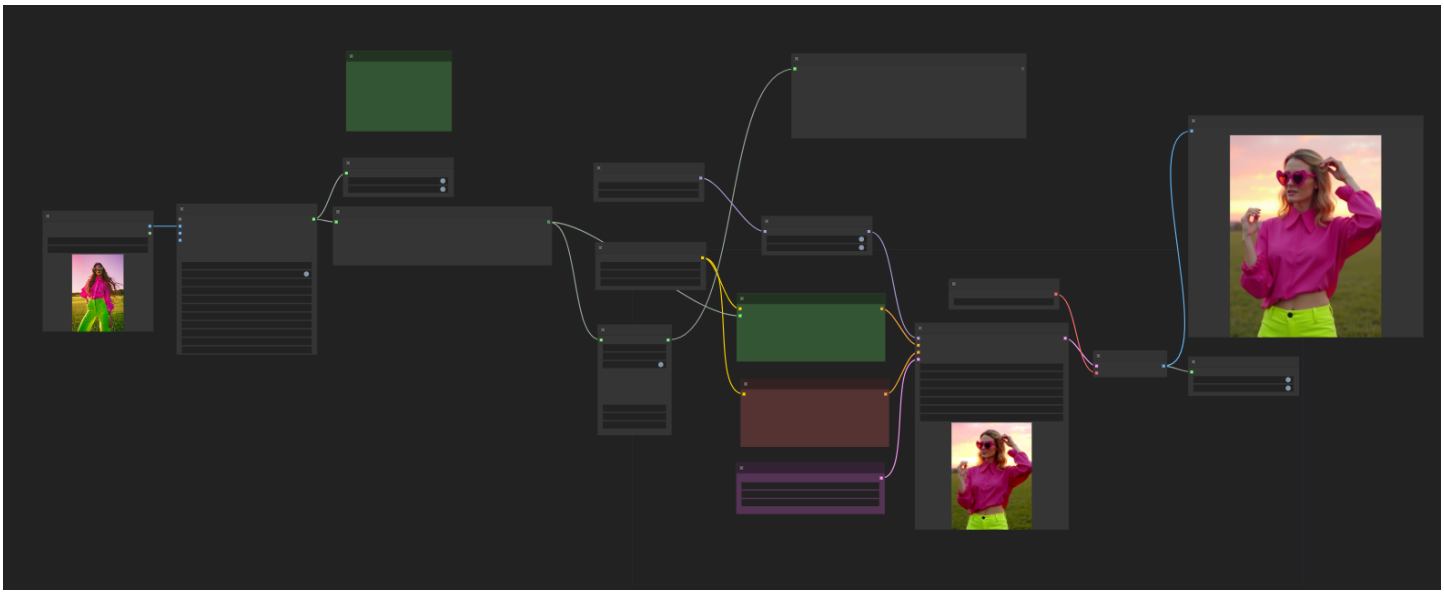
基础工作流验证

我们可以据此来做一个反推-生图工作流：

图反推生图工作流.json
15.60KB







有优化建议欢迎私我交流哈。

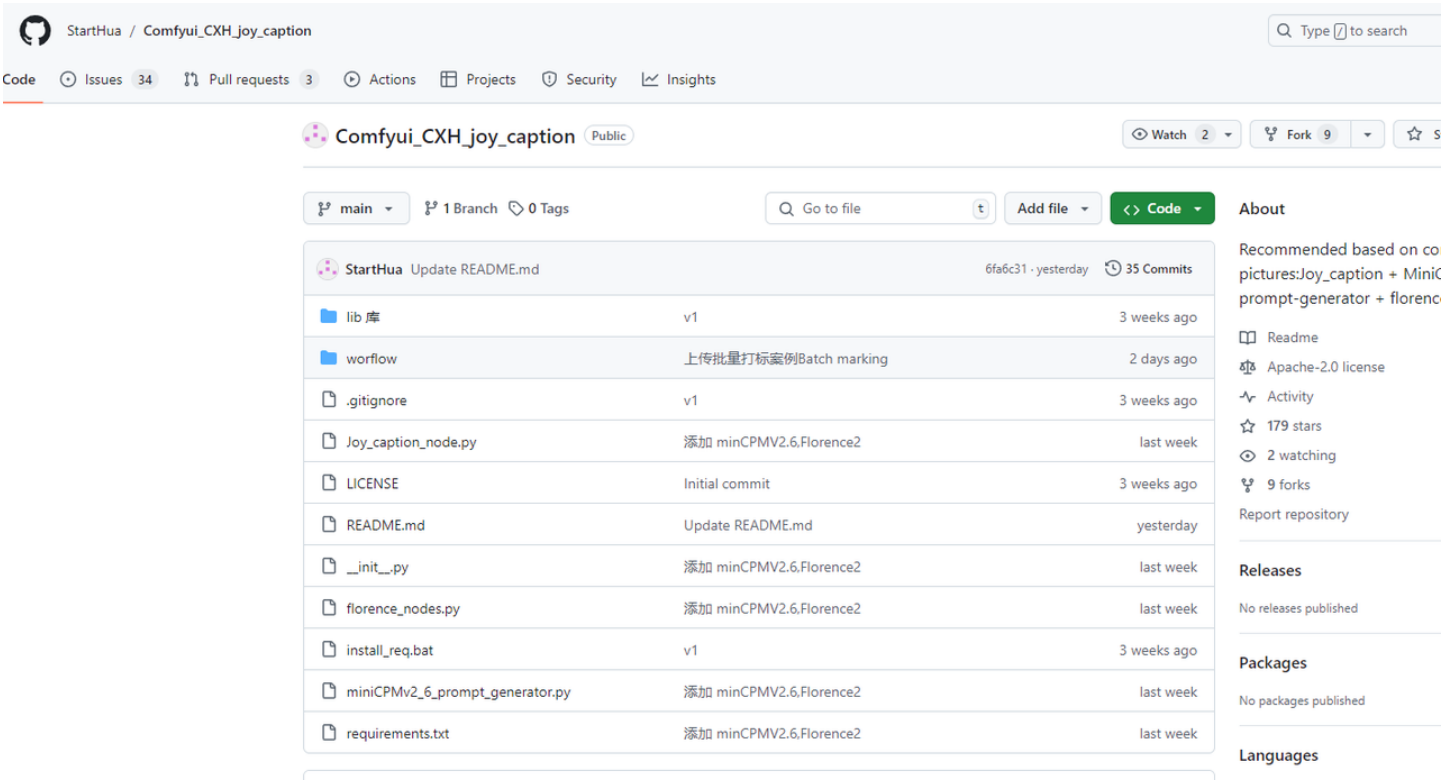
我们再来说一下Joy Caption

Joy Caption

Joy Caption目前星数较多的是 这个节点，但是比较难安装，劝退不少人。

1. 节点链接如下，可以依照介绍中所述安装节点、解决环境问题。

https://github.com/StartHua/Comfyui_CXH_joy_caption



Joy Caption 模型 (由 Fancy Feast 研发) 在 SigLIP 和 Llama3.1 的基础之上，使用 Adapter 模式，训练出更好的描述图像的模型，需要与 SigLIP 和 Llama3.1 混合使用，输入图像，输出一段语义丰富的图像描述。

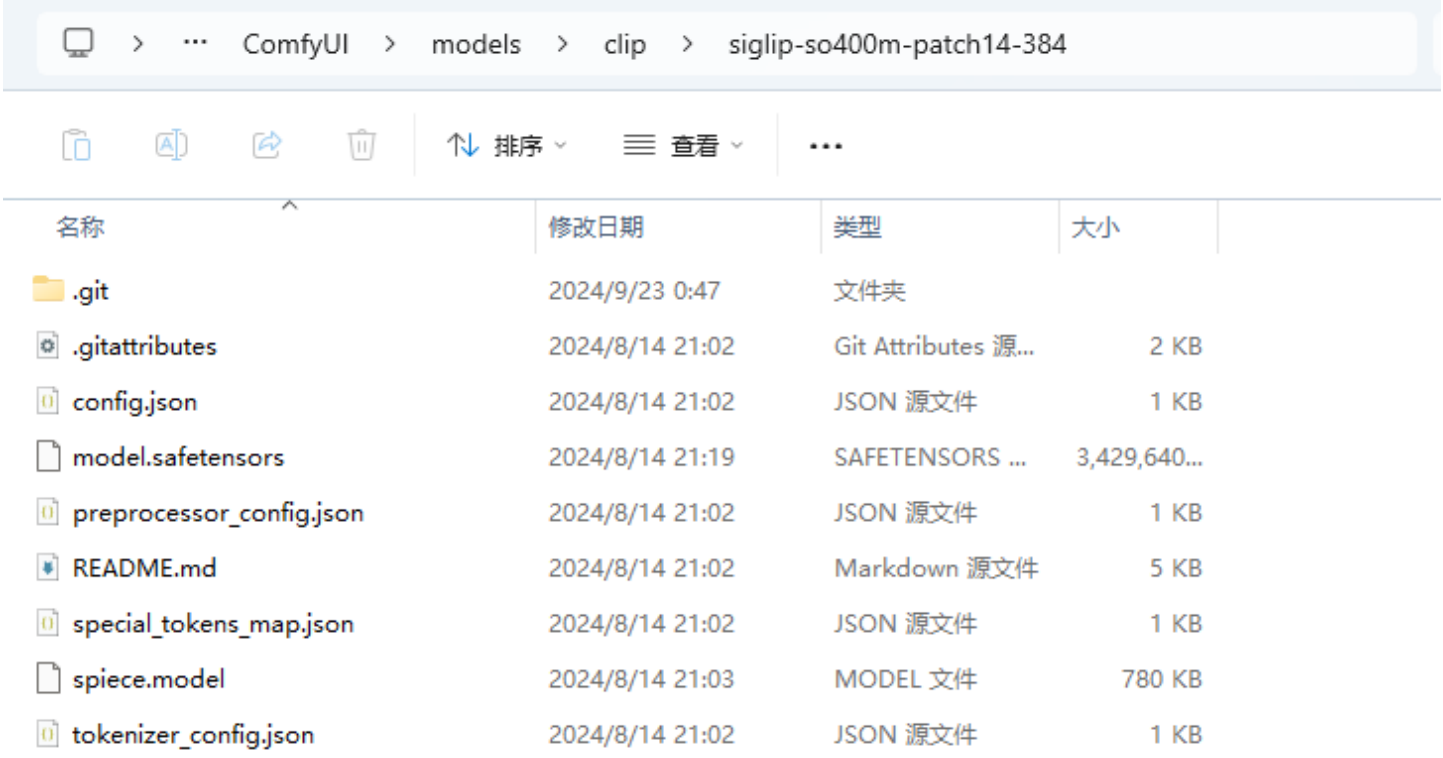
- Google 的 SigLIP (Sigmoid Loss for Language Image Pre-Training) 是一种改进的多模态模型，类似于 CLIP，但是采用了更优的损失函数。
- Meta-Llama-3.1-8B-bnb-4bit 是优化的多语言大语言模型，基于 Meta 的 Llama 3.1 架构，使用 BitsAndBytes 库进行 4-bit 量化，大幅减少内存使用，同时保持模型性能。

2. 所以我们下载模型需要上面介绍的：

三个模型（页面显示和实际大小有一定差异，正常，比如页面显示86MB的Joy_caption实际下载下来82MB）：

```
1   https://huggingface.co/google/siglip-so400m-patch14-384
```

3.5G，放到：你的\ComfyUI\models\clip\siglip-so400m-patch14-384

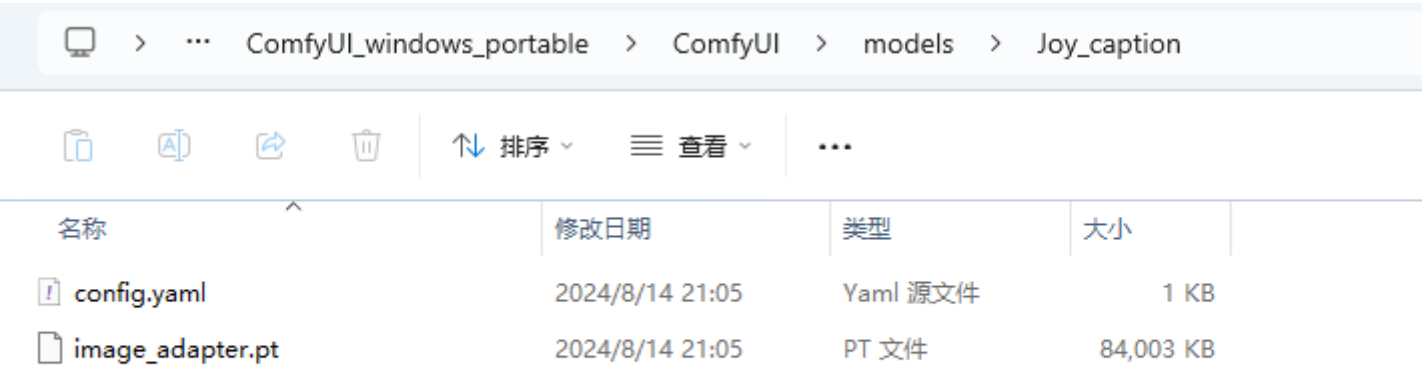



```
1   https://huggingface.co/unsloth/Meta-Llama-3.1-8B-bnb-4bit
```

5.7G，放到：你的\ComfyUI\models\LLM\Meta-Llama-3.1-8B-bnb-4bit

```
1  https://huggingface.co/spaces/fancyfeast/joy-caption-pre-
   alpha/tree/main/wpkklhc6
```

86MB，放到：你的\ComfyUI\models\Joy_caption



方便下载，这里我准备了网盘：

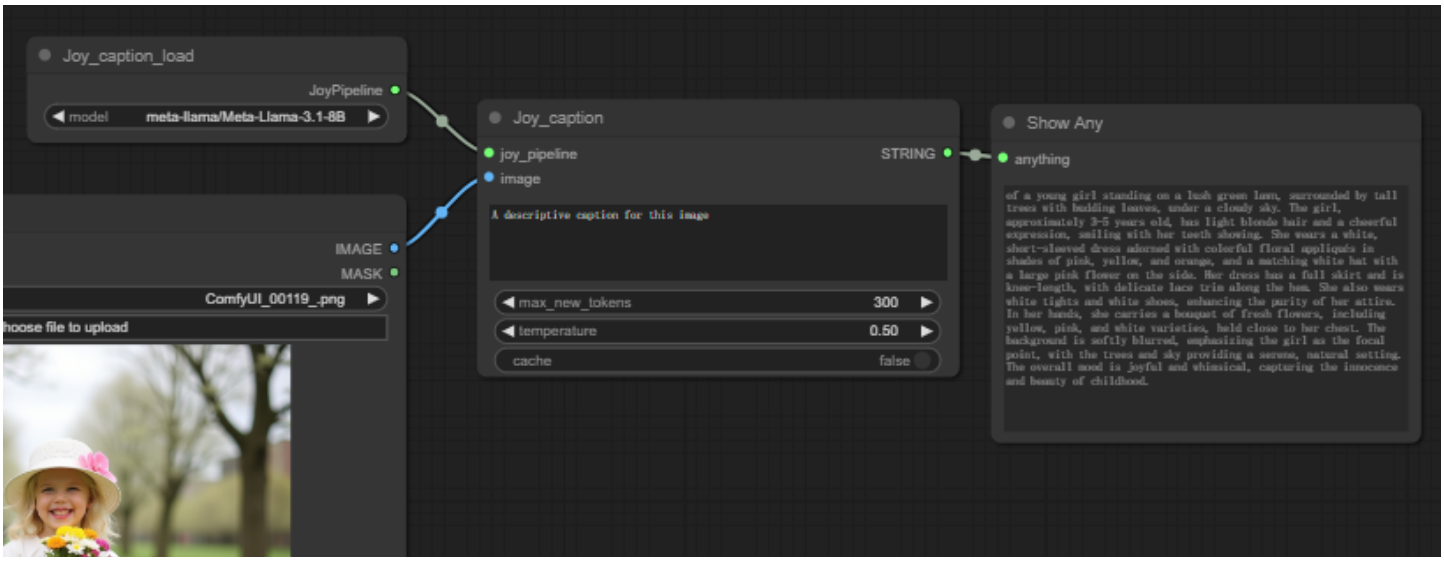
夸克网盘：链接：https://pan.quark.cn/s/eb7211b16c68

百度网盘：链接：https://pan.baidu.com/s/14IXKmN0QDbcoSNKKJMcmMA?pwd=pjyz

提取码：pjyz

3. 工作流验证：

joy.json
4.14KB



鉴于joy caption安装部署和环境问题总是难以实现，实在是不好处理，还有一个节点是：

这个节点也可以实现joy caption的功能

最终我们来类比一下florence2、MiniCPM-V、Joy Caption

其中有一个对比工作流

Min2.6+joy+Florence2.json
11.63KB

Min2_6
CXH_HG_Model_Load
model pzc163/MiniCPMv2_6-prompt-ge...
CXH_Min2_6_prompt_Run
A descriptive caption for this image
max_tokens 2048
temperature 0.70
seed 0
control_after_generate randomize
Show Any
The image presents a striking digital illustration of a knight in full armor, standing resolute against a backdrop of a tumultuous sky. The knight, clad in a dark, ornate suit of armor, wields a sword that glows with an otherworldly light, suggesting it may be infused with magical properties. The armor is intricately designed, featuring gold accents that catch the eye amidst the darker tones. The knight's helmet is adorned with a plume, adding to the regal appearance. The background is a dramatic canvas of dark clouds, hinting at an impending storm, which contrasts with the fiery glow emanating from the sword, creating a sense of tension and anticipation. The overall composition of the image suggests a narrative of conflict and heroism, with the knight poised to face whatever challenges lie ahead.
Load Image
image ComfyUI_temp_apaitm_00014_.png
choose file to upload
Joy_caption
Joy_caption_load
model unsloth/Meta-Llama-3.1-8B-bnb-4bit
Joy_caption
A descriptive caption for this image
max_new_tokens 2048
temperature 0.70
cache false
Show Any
1. This is a digital illustration depicting a majestic, armored warrior standing in a dramatic, stormy landscape. The warrior is a tall, imposing figure, clad in dark, metallic armor with intricate designs and sharp spikes. His helmet resembles a fearsome beast's head, with long, sharp horns curving backward. He wears a flowing cape that billows in the wind, adding a sense of movement and strength to his imposing stance. His eyes are hidden behind a visor, giving him an aura of mystery and intensity.
His left arm is sheathed in a long sword with a glowing, fiery blade, and his right hand grasps a similar sword with flames licking along its edge. The background is a tumultuous sky filled with dark clouds and flashes of lightning, creating a sense of impending danger and chaos. The ground is rugged and rocky, with small explosions of fiery orange gas rising from the surface, enhancing the sense of volatility and power.
The overall mood of the image is dark and foreboding, with a sense of otherworldly majesty. The artwork utilizes a detailed, realistic style, with a focus on the textures of the armor and the dynamic, swirling motion of the cape and the fiery blades.
florence2
CXH_DownloadAndLoadFlorence2Model
model thwri/CogFlorence-2.2-Large
precision fp16
attention sdpa
CXH_Florence2Run
text_input
task more_detailed_caption
fill_mask true
keep_model_loaded false
max_new_tokens 2048
num_beams 3
do_sample true
output_mask_select
seed 0
control_after_generate randomize
Show Any
A dramatic portrayal of a dark, armored warrior in a dynamic pose, wielding a long, fiery sword. The warrior wears ornate, dark armor with intricate designs and a helmet featuring a crown-like visor. The background is a stormy sky filled with dark clouds, and the ground is covered in fiery orange and yellow hues, indicating either ember dens or dusk. The overall color palette is dominated by dark blues, blacks, and fiery oranges, creating a sense of foreboding and intensity.