COMP7065 Innovative Laboratory

Spatial Control with ControlNet

After completion of this lab, students will be able to

1. Construct workflows with ControlNet nodes;
2. Apply various spatial control to output image by ControlNet, including:
   a) Human Pose,
   b) CannyEdge Maps,
   c) M-LSD Line Maps,
   d) HED Boundary Maps,
   e) Scribbles;
3. Integrate spatial control to their AI artworks;

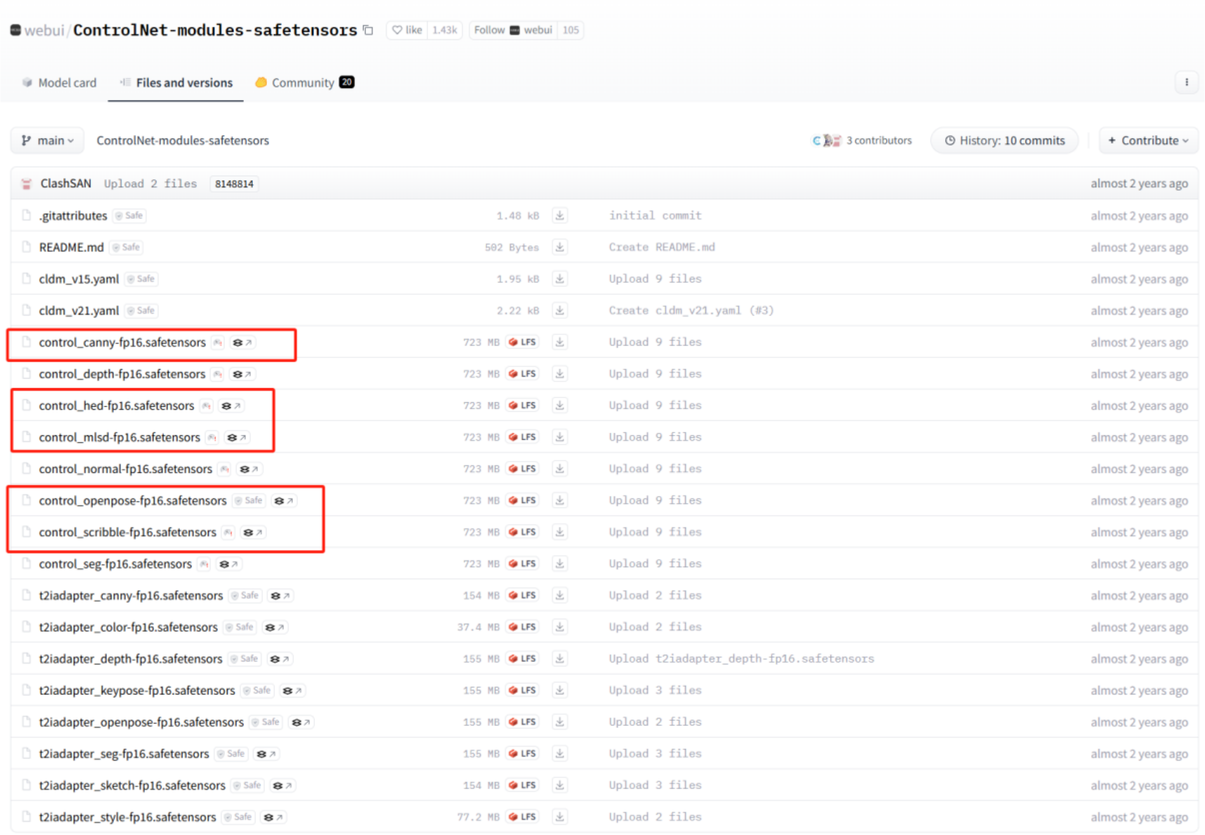
Table of Contents

1. Pose control with ControlNet 2
2. Explore different types of spatial control 9
3. Exercise 15

Credit to SHANG Sifeng

1. Pose control with ControlNet

Preparations: In this part, you will learn how to build a basic workflow that applies human pose control through **ControlNet** and **OpenPose**. Before we build up the workflow, you should download the model weights of ControlNet first. It is recommended to download the **pruned** versions on HuggingFace through the following link: https://huggingface.co/webui/ControlNet-modules-safetensors/tree/main

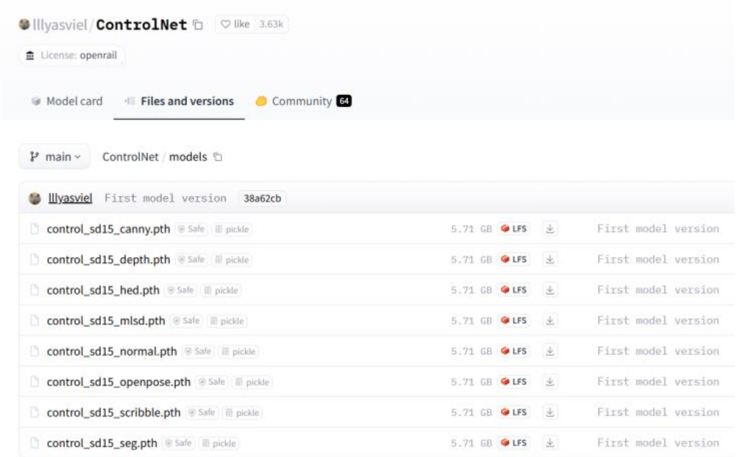


ClashSAN				Upload 2 files 8148814	almost 2 years ago
.gitattributes	Safe	1.48 kB		initial commit	almost 2 years ago
README.md	Safe	502 Bytes		Create README.md	almost 2 years ago
cldm_v15.yaml	Safe	1.95 kB		Upload 9 files	almost 2 years ago
cldm_v21.yaml	Safe	2.22 kB		Create cldm_v21.yaml (#3)	almost 2 years ago
control_canny-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_depth-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_hed-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_mlsd-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_normal-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_openpose-fp16.safetensors	Safe	723 MB	LFS	Upload 9 files	almost 2 years ago
control_scribble-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
control_seg-fp16.safetensors		723 MB	LFS	Upload 9 files	almost 2 years ago
t2iadapter_canny-fp16.safetensors	Safe	154 MB	LFS	Upload 2 files	almost 2 years ago
t2iadapter_color-fp16.safetensors	Safe	37.4 MB	LFS	Upload 2 files	almost 2 years ago
t2iadapter_depth-fp16.safetensors	Safe	155 MB	LFS	Upload t2iadapter_depth-fp16.safetensors	almost 2 years ago
t2iadapter_keypose-fp16.safetensors	Safe	155 MB	LFS	Upload 3 files	almost 2 years ago
t2iadapter_openpose-fp16.safetensors	Safe	155 MB	LFS	Upload 2 files	almost 2 years ago
t2iadapter_seg-fp16.safetensors	Safe	155 MB	LFS	Upload 3 files	almost 2 years ago
t2iadapter_sketch-fp16.safetensors	Safe	154 MB	LFS	Upload 3 files	almost 2 years ago
t2iadapter_style-fp16.safetensors	Safe	77.2 MB	LFS	Upload 2 files	almost 2 years ago

Please download the weights named **‘control_canny-fp16.safetensors’**, **‘control_hed-fp16.safetensors’**, **’control_mlsd-fp16.safetensors’**, **’control_openpose-fp16.safetensors’** and **‘control_scribble-fp16.safetensors’**, and place all of them into this folder: **‘ComfyUI\models\controlnet’**. Note that the postfix of the filename specifies the usage of the weights. For example, ‘_scribbles’ means that the ControlNet is trained to recognize spatial control from a scribble.

Credit to SHANG Sifeng

Alternatively, you can download the **full** model weights of ControlNet from its developer's model repository on HuggingFace: https://huggingface.co/lllyasviel/ControlNet/tree/main/models



lllyasviel / ControlNet

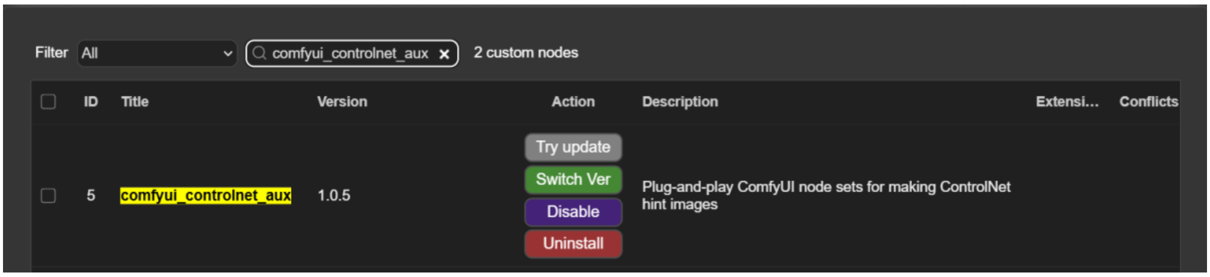
ControlNet / models

lllyasviel First model version 38a62cb

control_sd15_canny.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_depth.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_hed.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_mlsd.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_normal.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_openpose.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_scribble.pth	Safe	pickle	5.71 GB	LFS	First model version
control_sd15_seg.pth	Safe	pickle	5.71 GB	LFS	First model version

However, the full model weights occupy a considerable amount of disk storage (5.71GB for each type), and may potentially cause 'Cuda: out of memory' error during generation process. Meanwhile based on my own experience, the output quality doesn't have a significant improvement when replacing the pruned ControlNet with the full weights one. So it is more recommended that you directly use the pruned version of ControlNet.

Secondly, please install the custom node **'comfyUI_controlnet_aux'** by using **ComfyUI node manager**. This custom node provides you with a convenient method to apply image preprocessing, such as pose estimation and straight line detection, so as to cooperate with ControlNet spatial control.

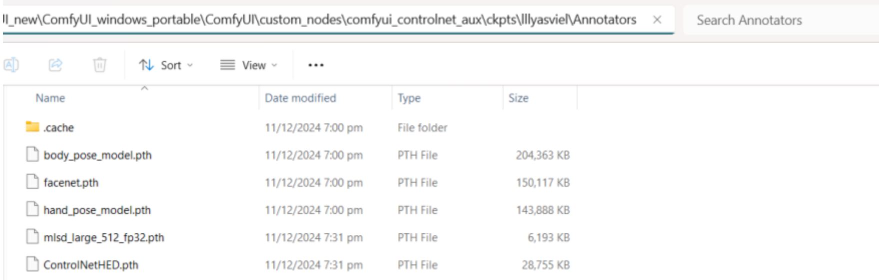


Filter: All | comfyui_controlnet_aux | 2 custom nodes

□	ID	Title	Version	Action	Description	Extensi...	Conflicts
□	5	comfyui_controlnet_aux	1.0.5	Try update Switch Ver Disable Uninstall	Plug-and-play ComfyUI node sets for making ControlNet hint images		

Credit to SHANG Sifeng

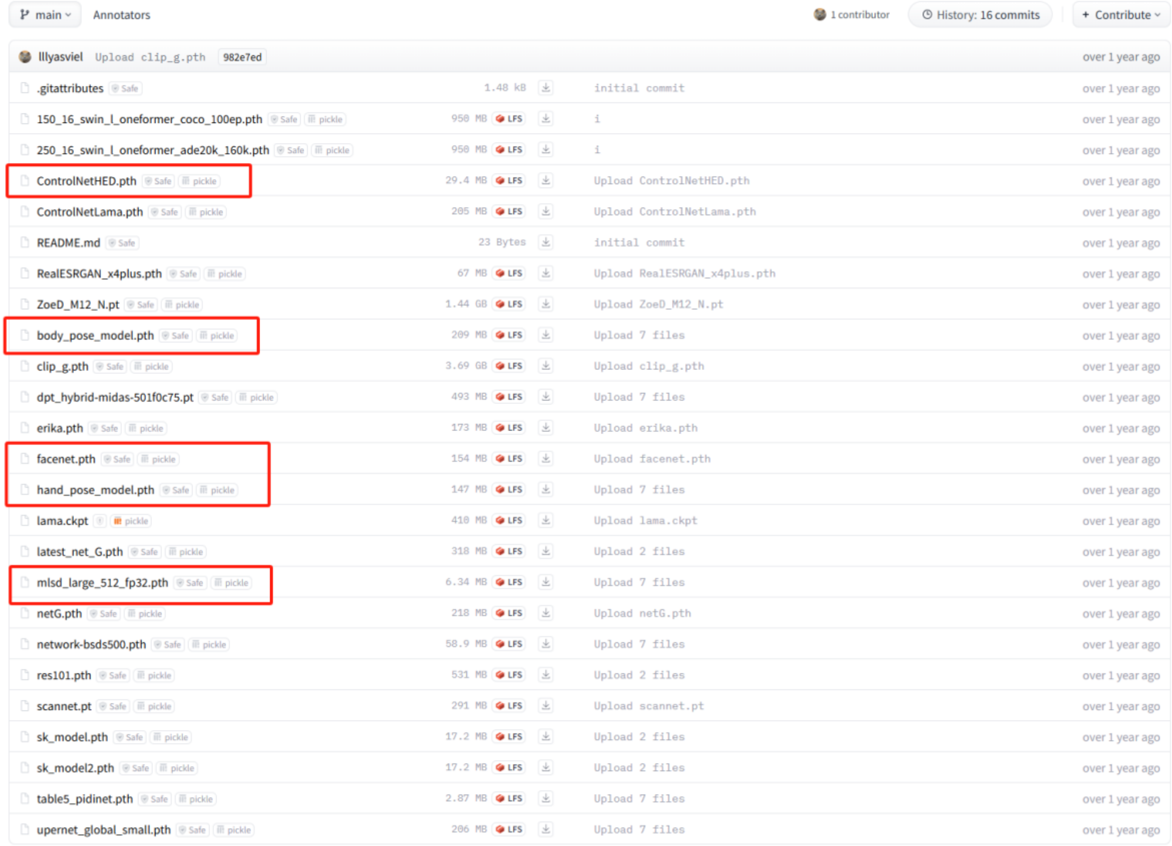
[Optional] During testing the showcases on our Lab's computer, I found that this custom node may not work smoothly in the latest ComfyUI. It sometimes may fail to automatically download the required model weights of the preprocessing network. To avoid this issue, you may need to manually download these weights from their official HuggingFace website by yourself, and place them into the folder: '**ComfyUI\custom_nodes\comfyui_controlnet_aux\ckpts\lllyasviel\Annotators**'.



ll_new\ComfyUI_windows_portable\ComfyUI\custom_nodes\comfyui_controlnet_aux\ckpts\lllyasviel\Annotators × | Search Annotators

Name	Date modified	Type	Size
.cache	11/12/2024 7:00 pm	File folder	
body_pose_model.pth	11/12/2024 7:00 pm	PTH File	204,363 KB
facenet.pth	11/12/2024 7:00 pm	PTH File	150,117 KB
hand_pose_model.pth	11/12/2024 7:00 pm	PTH File	143,888 KB
mlsd_large_512_fp32.pth	11/12/2024 7:31 pm	PTH File	6,193 KB
ControlNetHED.pth	11/12/2024 7:31 pm	PTH File	28,755 KB

The required model weights for this lecture's showcases are marked in the figure below.



lllyasviel Upload clip_g.pth 982e7ed over 1 year ago

.gitattributes	Safe	1.48 kB		initial commit	over 1 year ago
150_16_swin_l_oneformer_coco_100ep.pth	Safe pickle	950 MB	LFS	i	over 1 year ago
250_16_swin_l_oneformer_ade20k_160k.pth	Safe pickle	950 MB	LFS	i	over 1 year ago
ControlNetHED.pth	Safe pickle	29.4 MB	LFS	Upload ControlNetHED.pth	over 1 year ago
ControlNetLama.pth	Safe pickle	205 MB	LFS	Upload ControlNetLama.pth	over 1 year ago
README.md	Safe	23 Bytes		initial commit	over 1 year ago
RealESRGAN_x4plus.pth	Safe pickle	67 MB	LFS	Upload RealESRGAN_x4plus.pth	over 1 year ago
ZoeD_M12_N.pt	Safe pickle	1.44 GB	LFS	Upload ZoeD_M12_N.pt	over 1 year ago
body_pose_model.pth	Safe pickle	209 MB	LFS	Upload 7 files	over 1 year ago
clip_g.pth	Safe pickle	3.69 GB	LFS	Upload clip_g.pth	over 1 year ago
dpt_hybrid-midas-501f0c75.pt	Safe pickle	493 MB	LFS	Upload 7 files	over 1 year ago
erika.pth	Safe pickle	173 MB	LFS	Upload erika.pth	over 1 year ago
facenet.pth	Safe pickle	154 MB	LFS	Upload facenet.pth	over 1 year ago
hand_pose_model.pth	Safe pickle	147 MB	LFS	Upload 7 files	over 1 year ago
lama.ckpt	pickle	410 MB	LFS	Upload lama.ckpt	over 1 year ago
latest_net_G.pth	Safe pickle	318 MB	LFS	Upload 2 files	over 1 year ago
mlsd_large_512_fp32.pth	Safe pickle	6.34 MB	LFS	Upload 7 files	over 1 year ago
netG.pth	Safe pickle	218 MB	LFS	Upload netG.pth	over 1 year ago
network-bsds500.pth	Safe pickle	58.9 MB	LFS	Upload 7 files	over 1 year ago
res101.pth	Safe pickle	531 MB	LFS	Upload 2 files	over 1 year ago
scannet.pt	Safe pickle	291 MB	LFS	Upload scannet.pt	over 1 year ago
sk_model.pth	Safe pickle	17.2 MB	LFS	Upload 2 files	over 1 year ago
sk_model2.pth	Safe pickle	17.2 MB	LFS	Upload 2 files	over 1 year ago
table5_pidinet.pth	Safe pickle	2.87 MB	LFS	Upload 7 files	over 1 year ago
upernet_global_small.pth	Safe pickle	206 MB	LFS	Upload 7 files	over 1 year ago

Credit to SHANG Sifeng

Showcase: Human pose control is a basic function of ControlNet. Below is a showcase workflow adapted from an official example provided by RunComfy. You can download the source image (human.png) from this link.

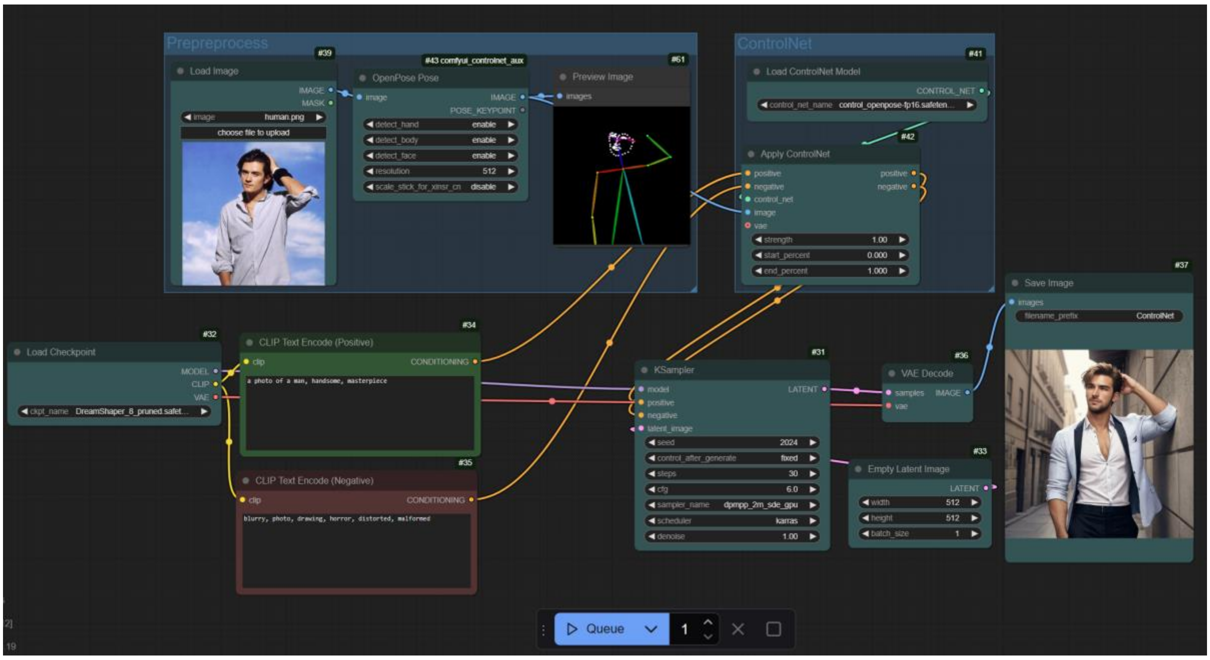


image: human.png
choose file to upload

detect_hand: enable
detect_body: enable
detect_face: enable
resolution: 512
scale_stick_for_xinsr_cn: disable

control_net_name: control_openpose-fp16.safeten...

strength: 1.00
start_percent: 0.000
end_percent: 1.000

ckpt_name: DreamShaper_8_pruned.safet...

a photo of a man, handsome, masterpiece

blurry, photo, drawing, horror, distorted, malformed

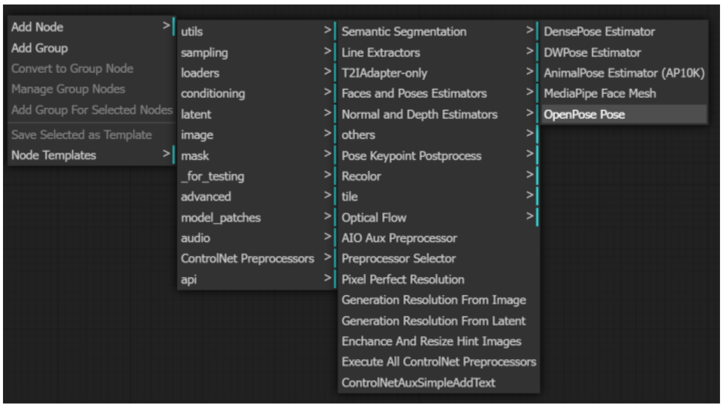
seed: 2024
control_after_generate: fixed
steps: 30
cfg: 6.0
sampler_name: dpmpp_2m_sde_gpu
scheduler: karras
denoise: 1.00

width: 512
height: 512
batch_size: 1

filename_prefix: ControlNet

Basically, this workflow in the showcase is equivalent to **adding two additional groups of nodes to the default image-to-text workflow**. The first group of nodes is responsible for **image preprocessing**, while the other group **configures the ControlNet**.

Image Preprocessing: In this showcase, the image preprocessing is to detect the keypoints of the human body such as face and joints using the OpenPose network. You should pay attention to the **'OpenPose Pose'** node, which is the core during image preprocessing. You can add it into your workflow by left clicking your mouse first and sequentially select 'Add Node -> ControlNet Preprocessors -> Faces and Poses Estimators -> OpenPose Pose'.



Credit to SHANG Sifeng

The functionality of this node is to apply pose estimation of a given figure photo or drawing. As illustrated by the Preview Image node, the output image consists of keypoints of the figure, and the keypoint image is then fed to ControlNet to apply spatial control to stable diffusion models. Description of each setting of the 'OpenPose Pose' node are presented in the table below. You may also refer to this link for more detailed explanations.

Setting	Description
detect_hand	When enabled, the OpenPose Pose node will detect keypoints of hands.
detect_body	When enabled, the OpenPose Pose node will detect keypoints of body parts.
detect_face	When enabled, the OpenPose Pose node will detect keypoints on the faces.
resolution	This parameter determines the size of keypoint image, which will always be in a square shape. It doesn't affect the resolution of the final output.
scale_stick_for_xinsr_cn	When enabled, the body lines will be plotted thicker and the result becomes more stable. See this post for more details: https://github.com/Fannovel16/comfyui_controlnet_aux/issues/447

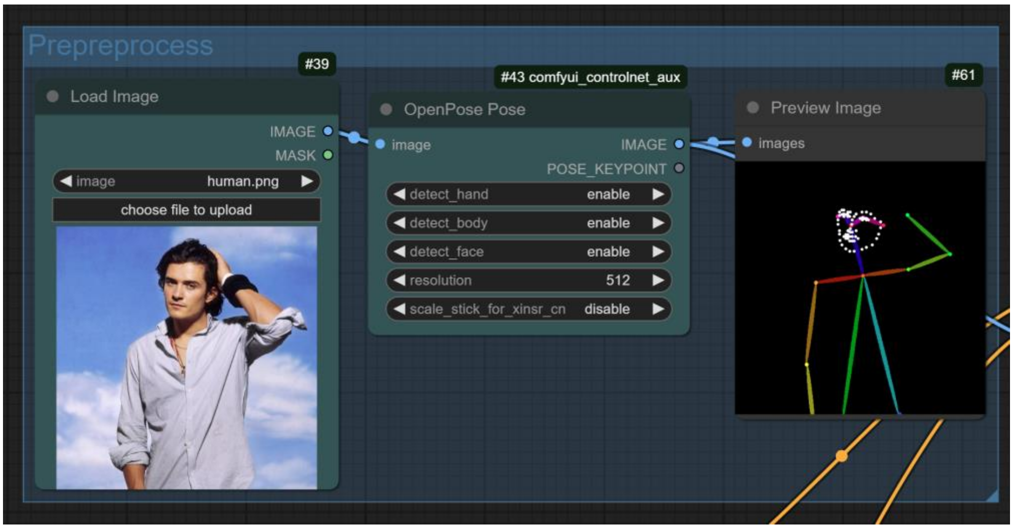
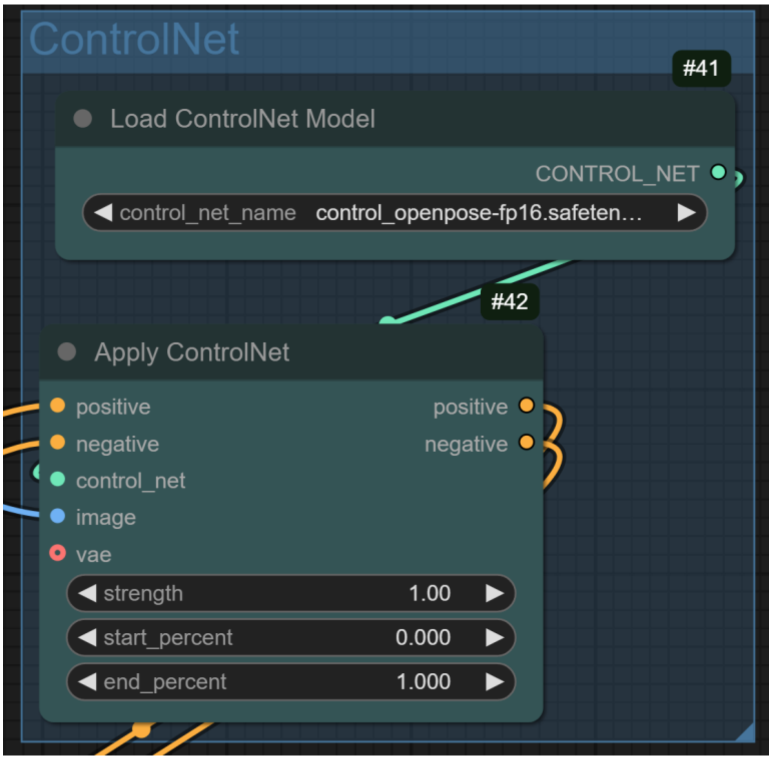


image human.png
detect_hand enable
detect_body enable
detect_face enable
resolution 512
scale_stick_for_xinsr_cn disable

Credit to SHANG Sifeng



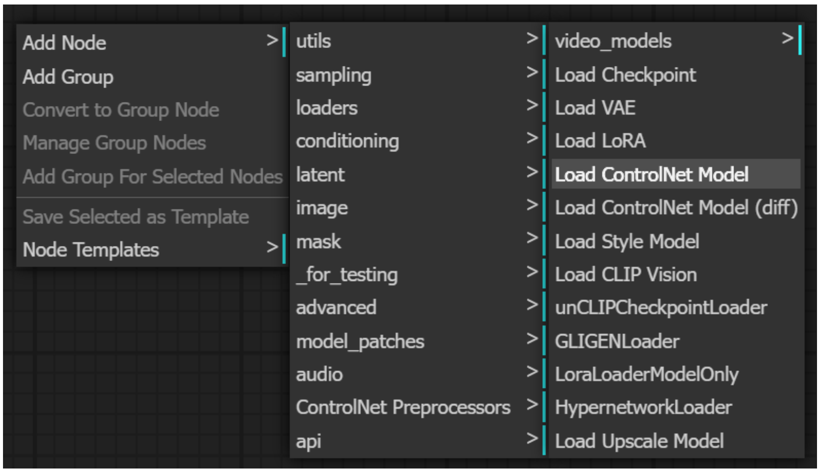
control_net_name control_openpose-fp16.safeten…

strength 1.00

start_percent 0.000

end_percent 1.000

Configuring the ControlNet: The other group of nodes, including **'Load ControlNet Model'** and **'Apply ControlNet'**, are responsible for configuring the ControlNet model. The functionality of **'Load ControlNet Model'** is easy to understand. It just loads the weights of a ControlNet model that you've downloaded during the preparation stage. Note that you should match the model weights with the type of spatial prompts. For exmample, in this showcase we utilize the keypoint image estimated by OpenPose as the prompt image for ControlNet, so the corresponding weights should be **'control_openpose-fp16.safetensors'**. One may not set the weights other than that, such as **'control_scribble_fp16.safetensors'**, which is intended for applying spatial prompts by scribbles. You may add the **'Load ControlNet Model'** node by sequentially selecting 'Add Node -> loaders -> Load ControlNet Model', or just type its name in the search bar popped after double clicking on the workspace.



The **‘Apply ControlNet’** node configures the ControlNet previously loaded by **‘Load ControlNet Model’** node. The description of each setting is presented in the table below. You can refer to this link for a detailed description on ComfyUI wiki.

Setting	Description
postivie	Positive conditioning features, such as the output of a positive CLIP text encoder.
negative	Negative conditioning features, such as the output of a negative CLIP text encoder.
contrl_net	The ControlNet model weights previously loaded
image	The preprocessed image for spatial constraint.
vae	Optional, determines whether ControlNet should affect the VAE models.
strength	A scalar that determines the strength of ControlNet’s effect on conditioning features.
start_percent	A float number between 0 and 1 that controls the start point of ControlNet’s effect during denoising steps.
end_percent	A float number between 0 and 1 that controls the end point of ControlNet’s effect during denoising steps.

It could be observed from the output image that the ControlNet is very effective. We didn’t describe anything about pose in our text prompt. But in the result, the man actually shares a very similar posture as the one in input.

You are strongly encouraged to rebuild this showcase workflow on your own as an exercise. Also try to adjust different settings and observe the impact on output images. This helps you to strengthen the understanding of the functionalities ControlNet model in an intuitive manner.

Credit to SHANG Sifeng

2. Explore different types of spatial control

Part 2 is just a simple showcase of how to apply spatial control of human pose through ControlNet. In Part 3, we will explore more types of spatial control by providing the ControlNet with different kinds of spatial prompts. Showcases in part 3 cover spatial prompts of Canny edge maps, Hough line maps, HED boundary maps and scibbles. It's also noteworthy that workflows in Part 3 are highly similar to the one in Part 2. The key differences lie in the **preprocessing node** and the corresponding **model weights** of ControlNet. They are adjusted according to the type of spatial prompts. Source images used in these showcases can be accessed through this link on Github.

Canny edge maps

Below is a showcase workflow that uses a Canny edge map as spatial prompt for ControlNet. The Canny edge detection method is a classical approach in image processing. It detects the edges of a given image using convolutional operations with Canny operator. The Canny operator assigns each pixel with its gradient, and the absolute value of it is used to judge whether the pixel is an edge. The final output of Canny edge map is a binary image where the edges are highlighted, as shown in the figures below.

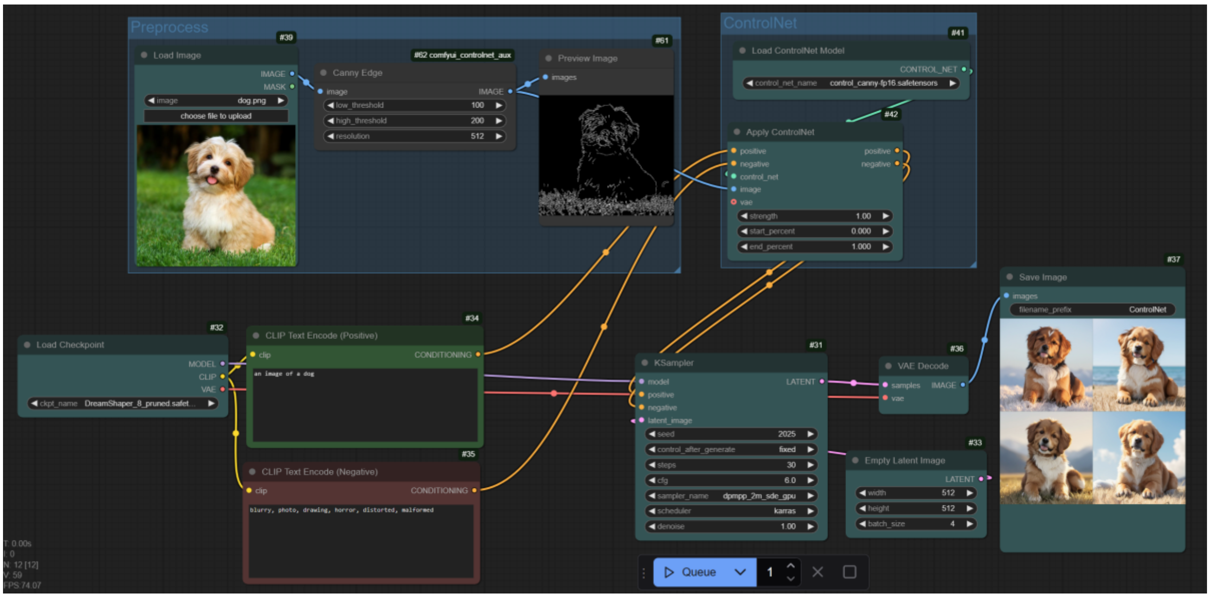


image dog.png

low_threshold 100

high_threshold 200

resolution 512

control_net_name control_canny-fp16.safetensors

strength 1.00

start_percent 0.000

end_percent 1.000

ckpt_name DreamShaper_8_pruned.safet...

```
an image of a dog
```

```
blurry, photo, drawing, horror, distorted, malformed
```

seed 2025

control_after_generate fixed

steps 30

cfg 6.0

sampler_name dpmpp_2m_sde_gpu

scheduler karras

denoise 1.00

width 512

height 512

batch_size 4

filename_prefix ControlNet

Credit to SHANG Sifeng

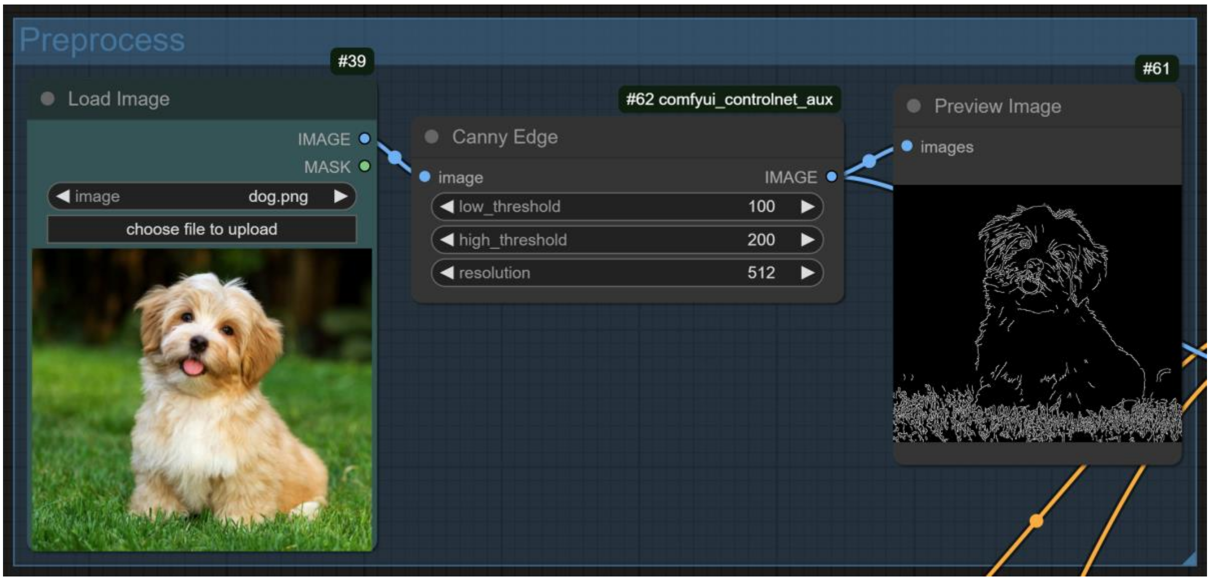


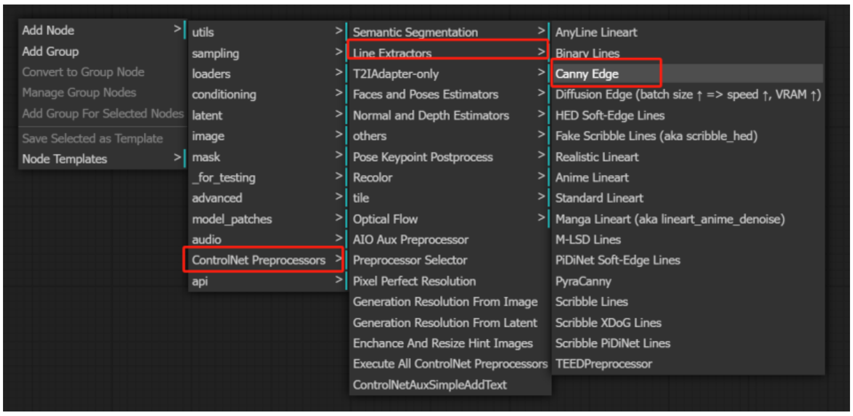
image dog.png

low_threshold 100

high_threshold 200

resolution 512

To apply Canny edge detection, you should modify the preprocess step by replacing **'OpenPose Pose'** node with the **'Canny Edge'** node, and set the model weights of ControlNet to **'control_canny-fp16.safetensors'**. You can find the node by clicking 'Add Node -> ControlNet Preprocessors -> Line Extractors -> Canny Edge'.



The settings of 'Canny Edge' node adjust the hyperparameter of the Canny edge detection algorithm. The description of each settings is presented in the table below. You can refer to this link for a more detailed explanation.

Setting	Description
low_threshold	An integer ranges from 0 to 255. Edge pixels whose gradient value is lower than this bound will be discarded. Descreasing this value result in more edge pixels.
high_threshold	An integer ranges from 0 to 255. Edge pixels whose gradient value is higher than this bound will be definitely preserved. Increasing this value result in less edge pixels.
resolution	This parameter determines the size of edge map, which will always be in a squared shape. It doesn't affect the resolution of the final output.

Credit to SHANG Sifeng

M-LSD line maps

M-LSD stands for mobile line segment detection. It is a deep learning-based, end-to-end line detection approach. In M-LSD, a light-weighted nerual network model detects the straight line segments inside an image, and returns a binary image in which the estimated line segments are highlighted, as shown in the figures below. It's notable that M-LSD line maps are very powerful for scene designs because the output image maintains almost the same structures as the spatial prompt image. You can refer to the original paper for more details about M-LSD method.

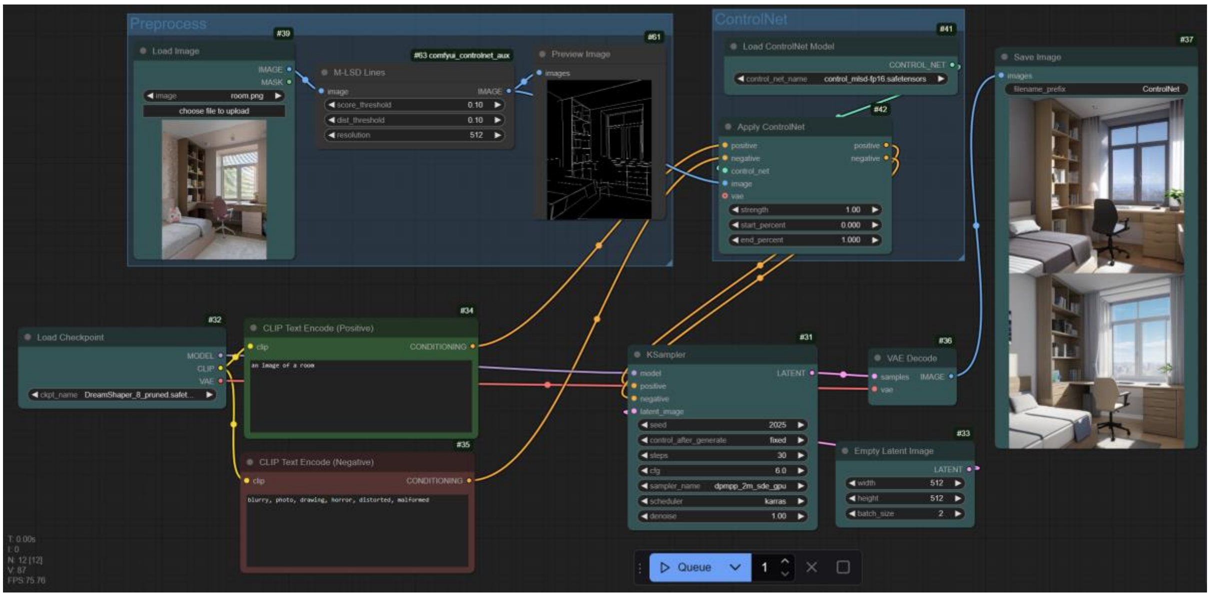


image: room.png
score_threshold 0.10
dist_threshold 0.10
resolution 512
control_net_name control_mlsd-fp16.safetensors
strength 1.00
start_percent 0.000
end_percent 1.000
filename_prefix ControlNet
ckpt_name DreamShaper_8_pruned.safet...

```
an image of a room
```

```
blurry, photo, drawing, horror, distorted, malformed
```

seed 2025
control_after_generate fixed
steps 30
cfg 6.0
sampler_name dpmpp_2m_sde_gpu
scheduler karras
denoise 1.00
width 512
height 512
batch_size 2

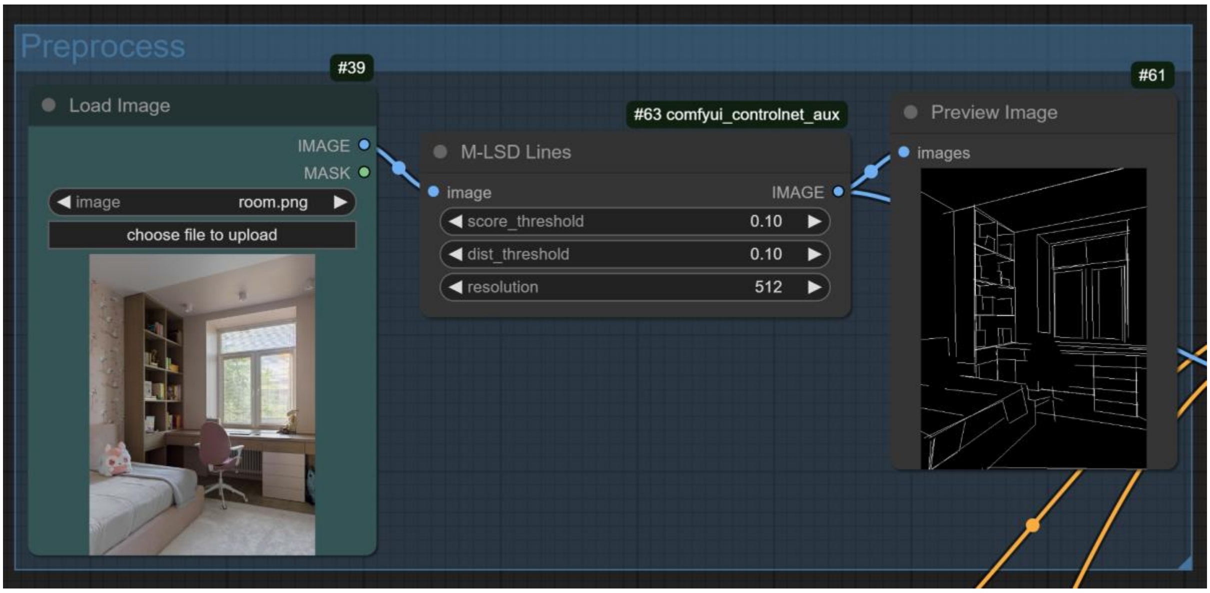
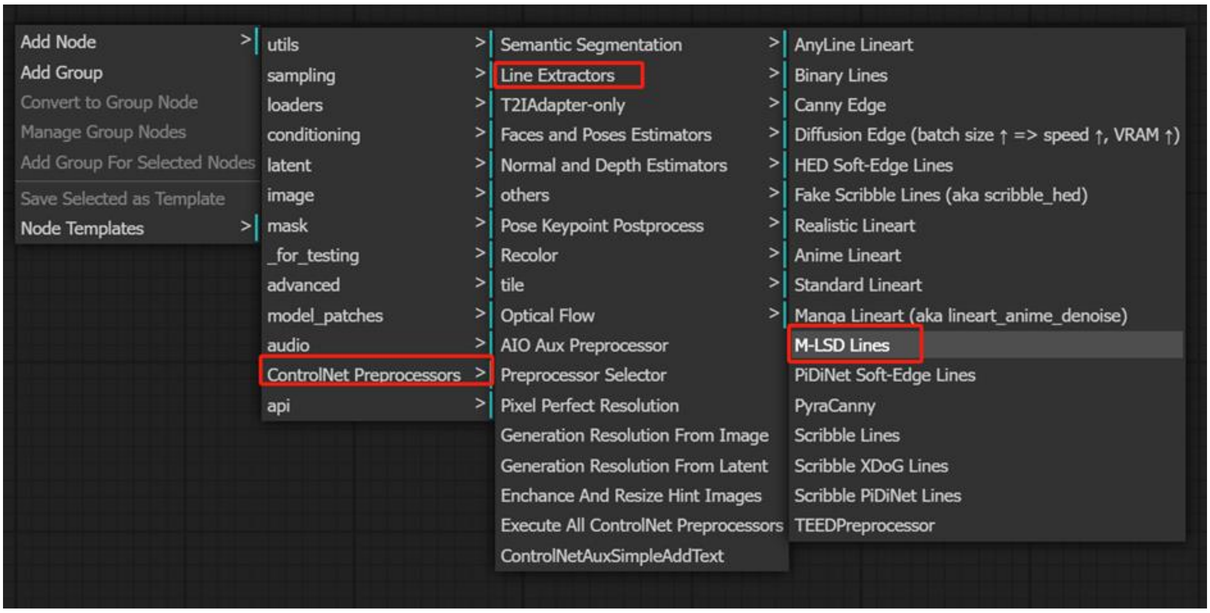


image room.png
score_threshold 0.10
dist_threshold 0.10
resolution 512

Credit to SHANG Sifeng

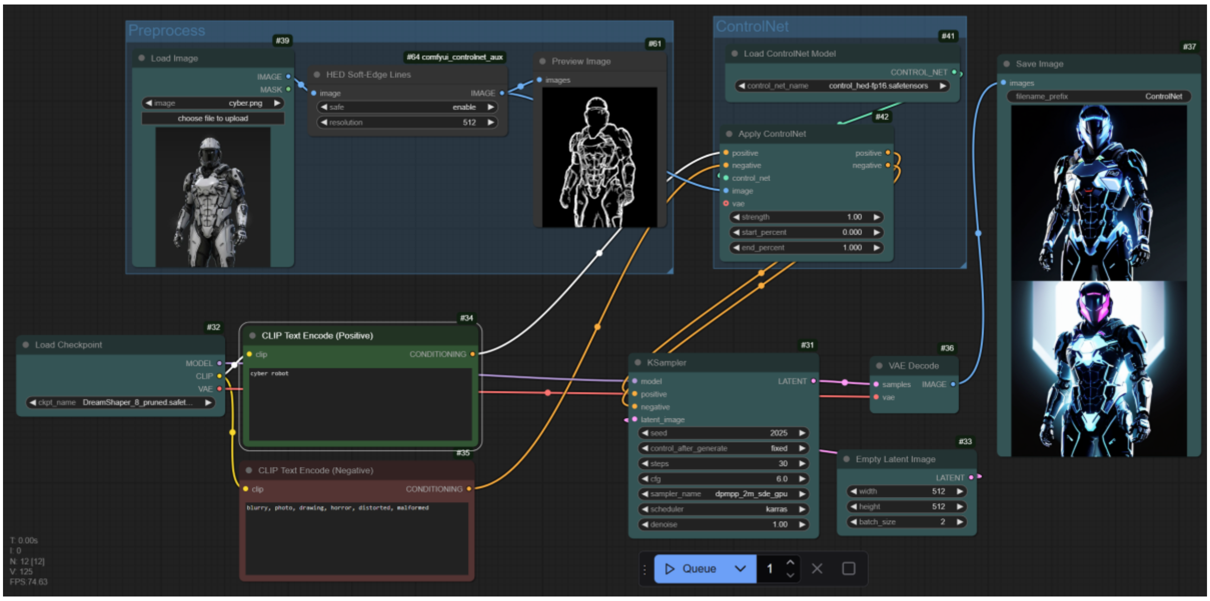
To apply M-LSD method, you should use the **‘M-LSD Lines’** in preprocess stage, and change the weights of ControlNet to **‘control_mlsd-fp16.safetensors’**. The figure below shows how to find the node, and the table below explains the settings of it. You can read the descriptions provided by this link for more details.



Setting	Description
score_threshold	A float number ranges from 0.01 to 2.0. It determines the minimum confidence score for a line to be valid. Descreasing this value result in more lines, but will potentially add noises.
dist_threshold	A float number ranges from 0.01 to 20.0. It determines the maximum allowable distance beteen points for them to be considered part of the same line.
resolution	This parameter determines the size of line map, which will always be in a squared shape. It doesn’t affect the resolution of the final output.

HED boundary maps

HED stands for holistically-nested edge detection. Similar to M-LSD, it is also a deep-learning based, end-to-end method. HED performs fast prediction on the edges of an given image using a fully convolutional nerual network. HED boundaries are different with the Canny edges, because they are ‘soft’ edge maps instead of binary. This means that HED boundaries are more detailed. If you are interested in technical details of this method, you can refer to its original paper on arxiv.



Preprocess #39
Load Image
IMAGE
MASK
image cyber.png
choose file to upload

#64 comfyui_controlnet_aux
HED Soft-Edge Lines
image IMAGE
safe enable
resolution 512

#61 Preview Image
images

ControlNet #41
Load ControlNet Model
CONTROL_NET
control_net_name control_hed-fp16.safetensors

#42 Apply ControlNet
positive positive
negative negative
control_net
image
vae
strength 1.00
start_percent 0.000
end_percent 1.000

#37 Save Image
images
filename_prefix ControlNet

#32 Load Checkpoint
MODEL
CLIP
VAE
ckpt_name DreamShaper_8_pruned.safet...

#34 CLIP Text Encode (Positive)
clip CONDITIONING
cyber robot

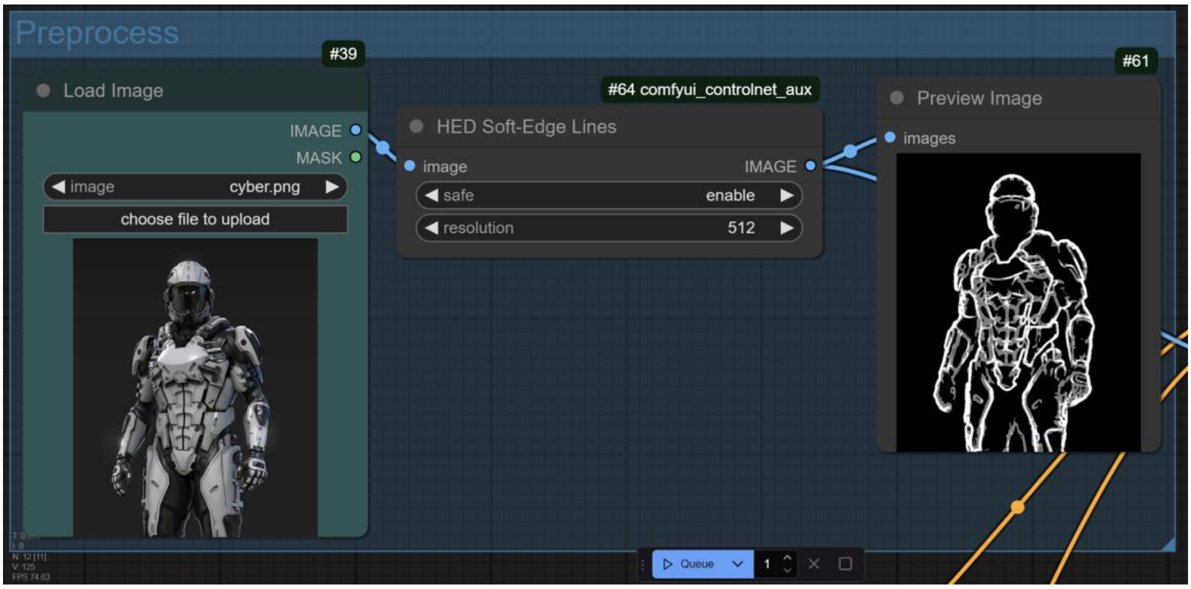
#35 CLIP Text Encode (Negative)
clip CONDITIONING
blurry, photo, drawing, horror, distorted, malformed

#31 KSampler
model LATENT
positive
negative
latent_image
seed 2025
control_after_generate fixed
steps 30
cfg 6.0
sampler_name dpmpp_2m_sde_gpu
scheduler karras
denoise 1.00

#36 VAE Decode
samples IMAGE
vae

#33 Empty Latent Image
LATENT
width 512
height 512
batch_size 2

T: 0.00s
I: 0
N: 12 [12]
V: 125
FPS:74.63



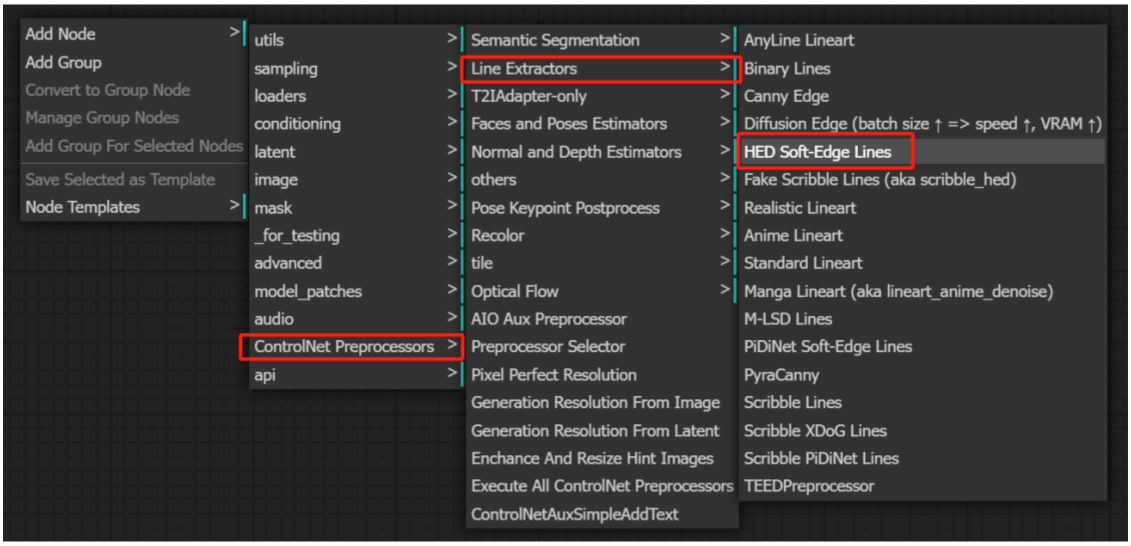
Preprocess #39
Load Image
IMAGE
MASK
image cyber.png
choose file to upload

#64 comfyui_controlnet_aux
HED Soft-Edge Lines
image IMAGE
safe enable
resolution 512

#61 Preview Image
images

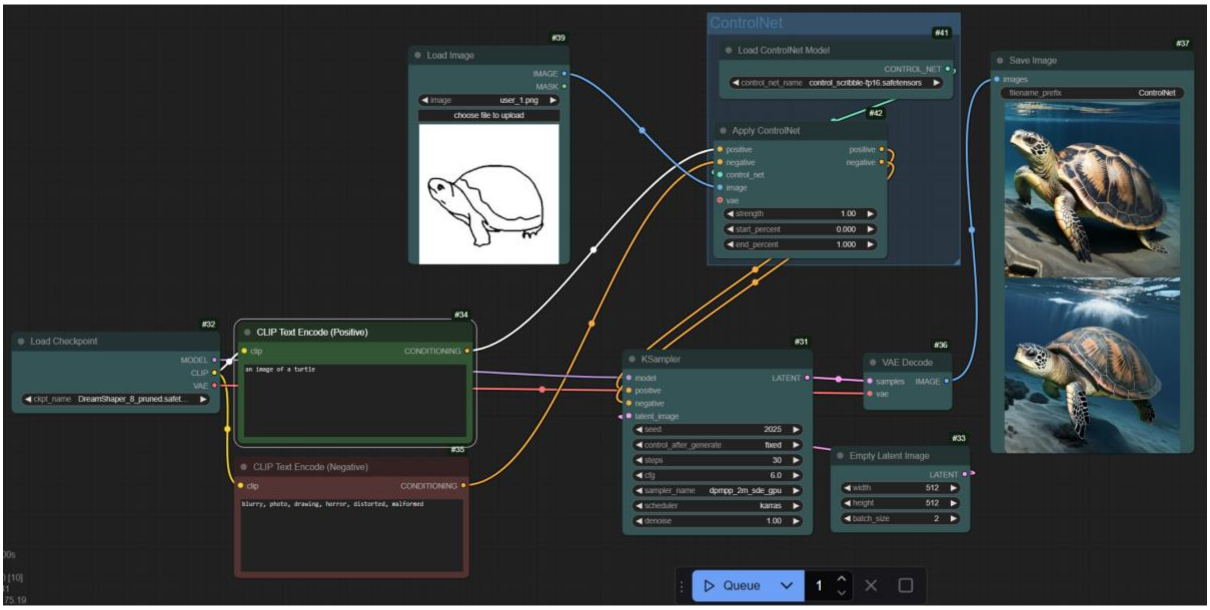
To apply the HED method to ControlNet spatial constraint, you should use the **‘HED Soft-Edge Lines’** node in the preprocess stage, and select the **‘control_hed-fp16.safetensors’** as weights for ControlNet. To find the node, you should sequentially click on ‘Add Node -> ControlNet Preprocessors -> Line Exactors -> HED Soft-Edge Maps’, as shown in the figure below. The setting ‘safe’ determines if the additional safety checks should be toggled. You can read the detailed explanation provided by this link for more instructions.

Credit to SHANG Sifeng



Scribbles

The last showcase is to provide a scribble as spatial prompt. Compared with others, this may be the simplest one since no preprocess is required, and you only need change the model wights of ControlNet to **‘controlnet_scribble-fp16.safetensors’** and feed it with a scribble or skectch.



Load Image #39: image user_1.png

Load Checkpoint #32: ckpt_name DreamShaper_8_pruned.safet...

CLIP Text Encode (Positive) #34: an image of a turtle

CLIP Text Encode (Negative) #35: blurry, photo, drawing, horror, distorted, malformed

Load ControlNet Model #41: control_net_name control_scribble-fp16.safetensors

Apply ControlNet #42: strength 1.00, start_percent 0.000, end_percent 1.000

KSampler #31: seed 2025, control_after_generate fixed, steps 30, cfg 6.0, sampler_name dpmpp_2m_sde_gpu, scheduler karras, denoise 1.00

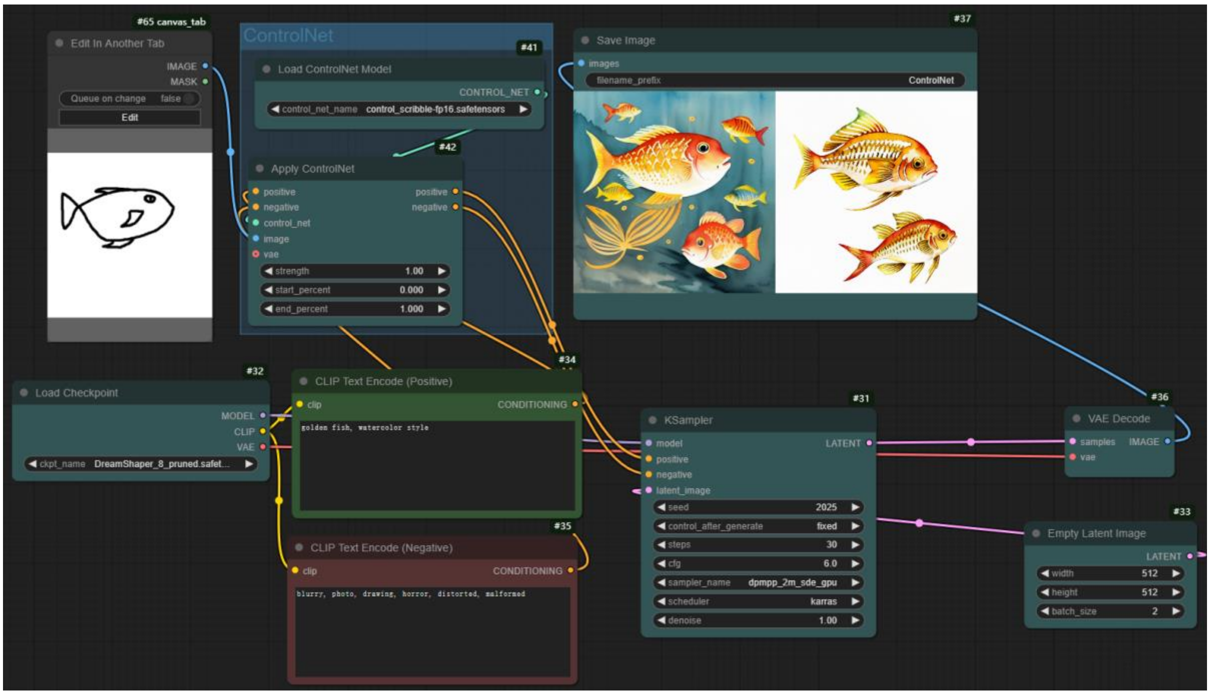
Empty Latent Image #33: width 512, height 512, batch_size 2

Save Image #37: filename_prefix ControlNet

Credit to SHANG Sifeng

3. Exercise

Draw a scribble by yourself using the ‘Canvas Tab’ node and feed it as a spatial prompt to ControNet. Feel free to design anything you like! If you forgot what is a ‘Canvas Tab’ node, please check Exercise (2) in Lab 7 for more details.



Queue on change false

control_net_name control_scribble-fp16.safetensors

strength 1.00

start_percent 0.000

end_percent 1.000

filename_prefix ControlNet

ckpt_name DreamShaper_8_pruned.safet...

golden fish, watercolor style

blurry, photo, drawing, horror, distorted, malformed

seed 2025

control_after_generate fixed

steps 30

cfg 6.0

sampler_name dpmpp_2m_sde_gpu

scheduler karras

denoise 1.00

width 512

height 512

batch_size 2

Final Submission: Please submit the workflow and 10 images generated from your drawings (along with the manual drawings), as a single zip file.

Credit to SHANG Sifeng