ControlNet: Customizing and Directing AI-Generated Art

Shaw Xiao

INTRODUCTION

In this final project I delved into the image generation using Stable Diffusion 1.5 and ControlNet models.

- **Stable Diffusion 1.5** known for its ability to generate detailed and coherent visuals based on prompts.
- **ControlNet** is a neural network structure that allows us to control diffusion models like Stable Diffusion by incorporating additional inputs, such as sketches, depth maps, or segmentation maps. This enables more precise manipulation of the generated images.

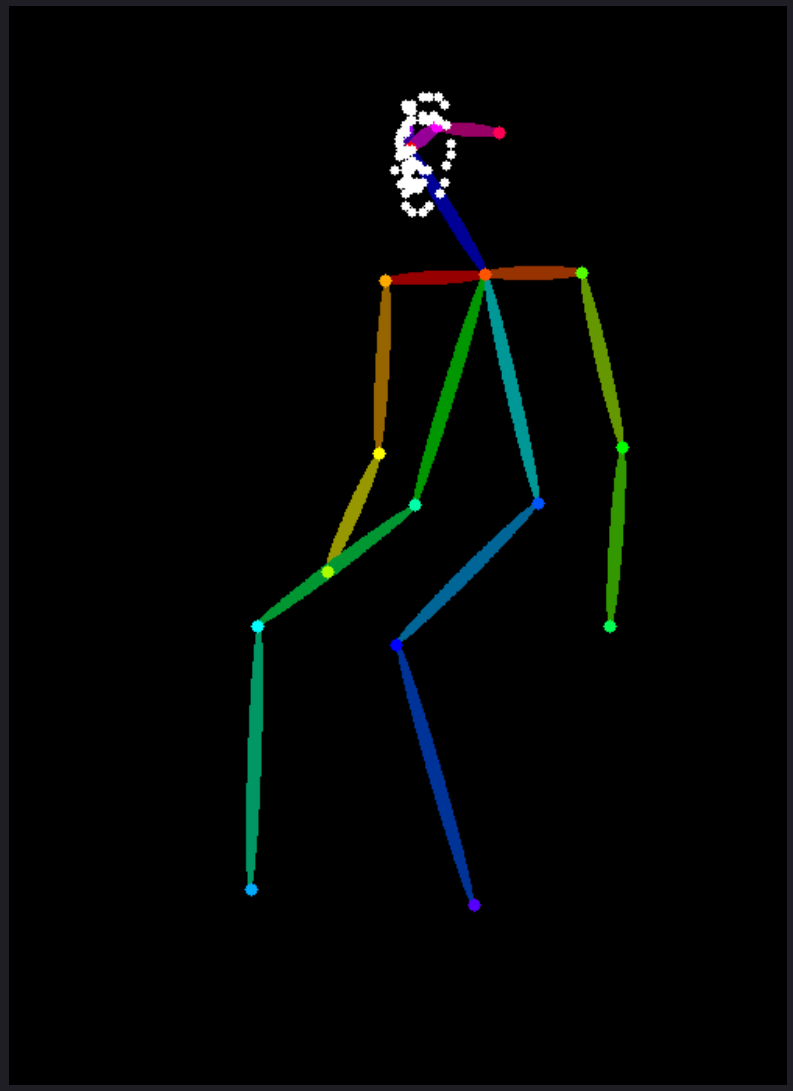
In this project I wanted to see how far I could go in customizing and directing the output of these models to produce images that are not only visually good but also reflective of specific artistic concepts. By using ControlNet, I aimed to overcome some limitations of traditional text-to-image generation, such as lack of control over specific elements in the scene.

ControlNet Models Overview

- **OpenPose:**
  - Controls human figure poses and movements
  - Ideal for character design and motion capture
- **Canny Edge Detection:**
  - Captures outlines and edges of images
  - Enhances structural control and shape definition
- **Depth:**
  - Adds depth and spatial perception
  - Transforms flat images into 3D-like scenes
- **Segmentation:**
  - Performs semantic segmentation of image regions
  - Allows distinct styling for different parts of an image

OpenPose Model

- Function: Used to control the pose and movement of human figures.
- Application: By inputting human skeleton images, we can generate images with specific poses, useful for character design, dance motion capture, and more.



OpenPose Model

ORINGINAL IMAGE

OpenPose Model Output

A 3D render of a girl sitting comfortably on a chair in a glass made house, in the art style of The Sims 4, highly detailed, smooth textures, vibrant colors, soft lighting, cheerful atmosphere, digital illustration
Negative prompt: blurry, low quality, deformed, disfigured, extra limbs, text, watermark, out of frame, distorted hands

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 7, Seed: 2159576166, Size: 710x1000, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Denoising strength: 0.75, Token merging ratio: 0.2, ControlNet 0: "Module: openpose_full, Model: control_v11p_sd15_openpose [cab727d4], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Noise multiplier: 1.05, Version: v1.9.3

OpenPose Model Output

Adjust prompt:

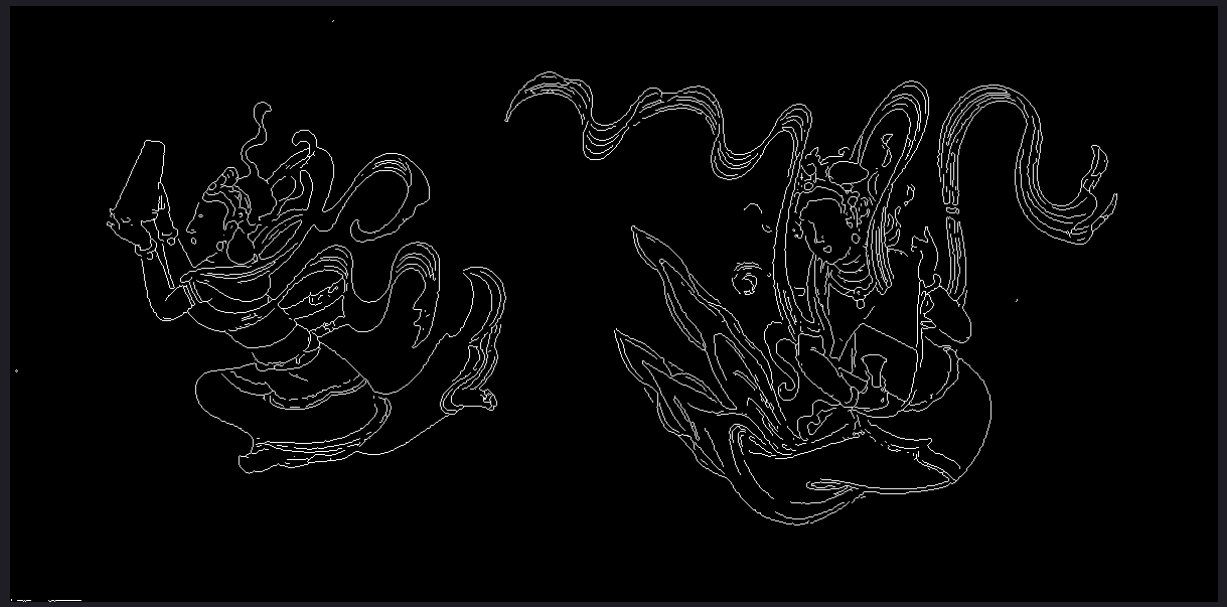
A 3D render of a girl sitting comfortably on a chair in a garden, in the art style of The Sims 4, highly detailed, smooth textures, vibrant colors, soft lighting, cheerful atmosphere, digital illustration

Output analysis:

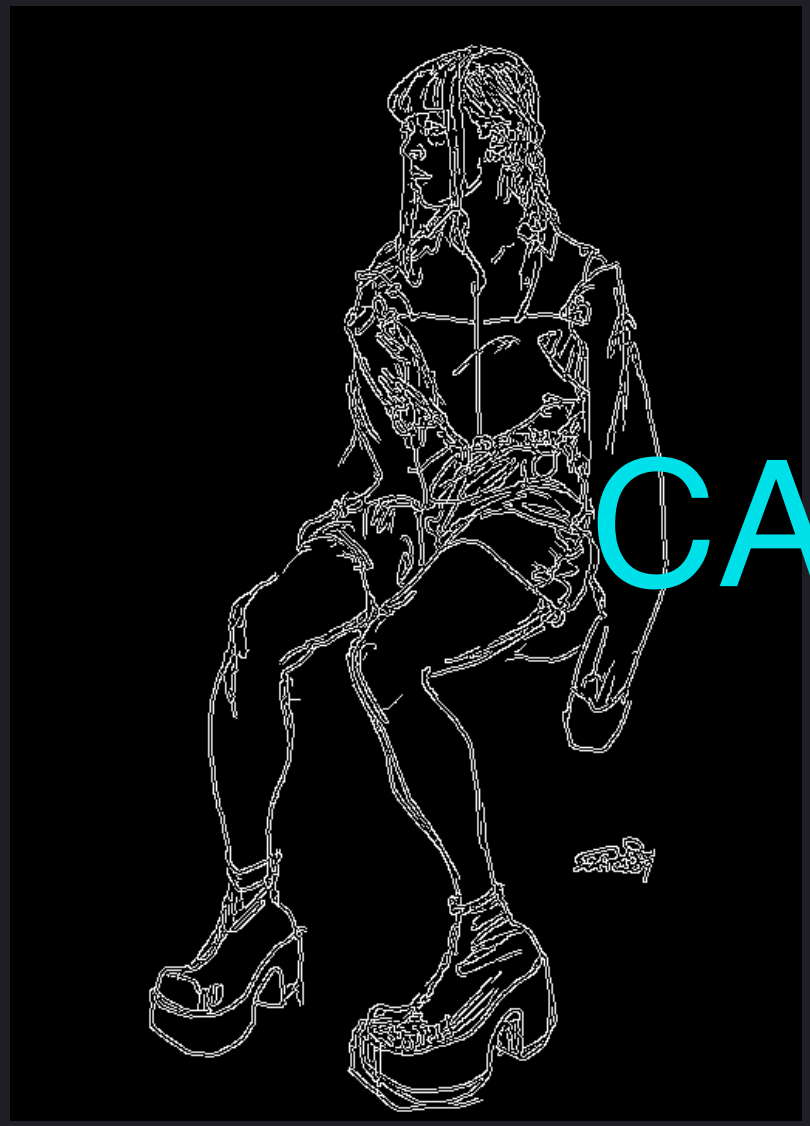
The generated image accurately replicates the actions and poses of the original input, demonstrating the effectiveness of the OpenPose model in controlling human figure movements. However, the background of the generated image appears somewhat disorganized and contains several blank spaces, which detracts from the overall visual coherence and richness of the scene. While the foreground maintains high detail and clarity, the lack of structured background elements makes the environment feel less immersive. To enhance future outputs, improving background organization and reducing empty areas by incorporating additional textures or complementary elements will be essential.

Canny Edge Detection Model

- Function: Captures the outlines and edges of images.
- Application: By inputting edge maps, we can control the overall shape and structure of the image, retaining the details of the original sketch.



Canny Edge Detection Model

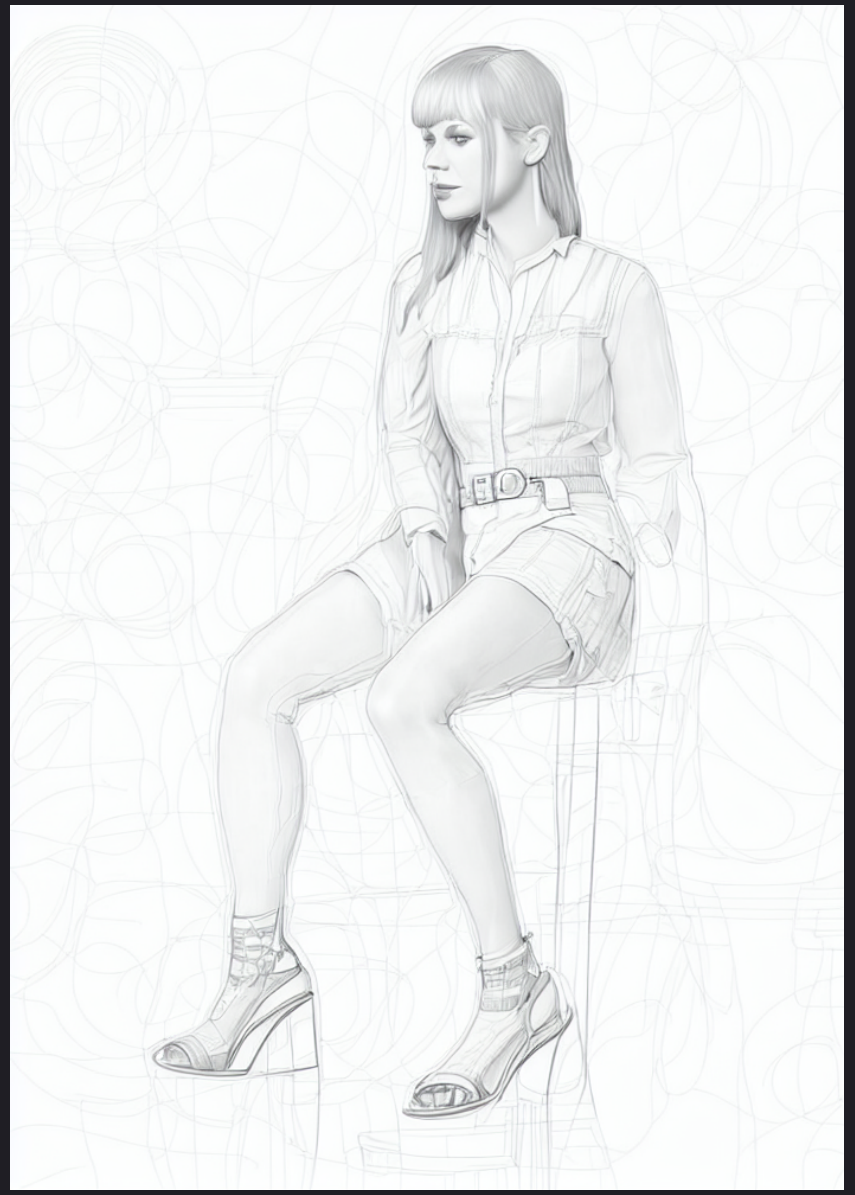


CANNY IMAGE

Canny Edge Detection Model Output

A 3D render of a girl sitting comfortably on a chair in a garden, in the art style of The Sims 4, highly detailed, smooth textures, vibrant colors, soft lighting, cheerful atmosphere, digital illustration
Negative prompt: blurry, low quality, deformed, disfigured, extra limbs, text, watermark, out of frame, distorted hands

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 7, Seed: 3231923029, Size: 710x1000, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Denoising strength: 0.75, Token merging ratio: 0.2, ControlNet 0: "Module: canny, Model: control_v11p_sd15_canny [d14c016b], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 100.0, Threshold B: 200.0, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: My prompt is more important", Noise multiplier: 1.05, Version: v1.9.3



Mode Comparison

My prompt is more important

Balance

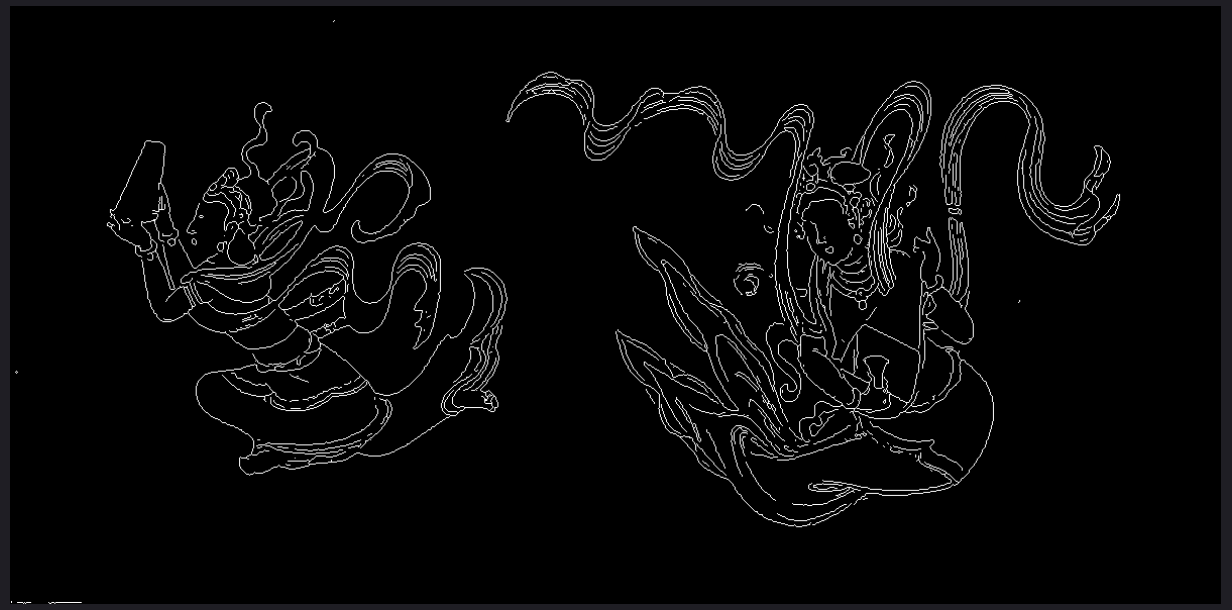
ControlNet is more important

Canny Edge Detection Model Output

A 3D render of a girl sitting comfortably on a chair in a garden, in the art style of The Sims 4, highly detailed, smooth textures, vibrant colors, soft lighting, cheerful atmosphere, digital illustration
Negative prompt: blurry, low quality, deformed, disfigured, extra limbs, text, watermark, out of frame, distorted hands

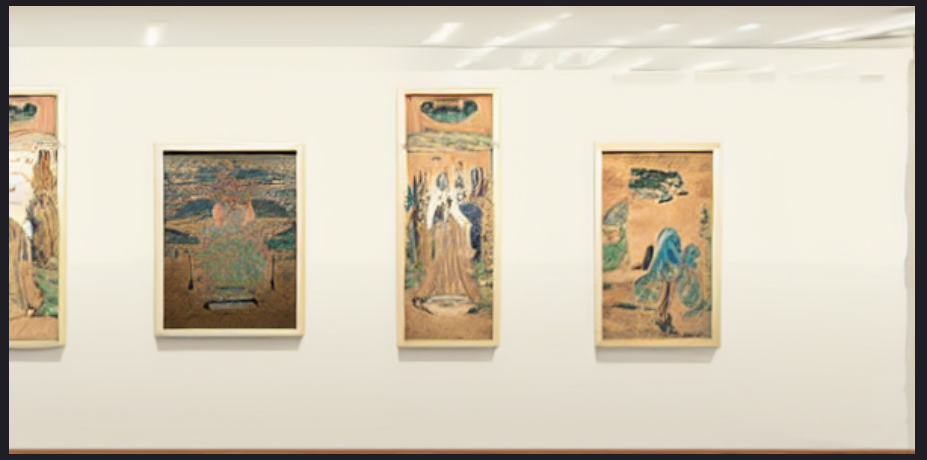
Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: **12**, Seed: 3231923029, Size: 710x1000, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Denoising strength: 0.75, Token merging ratio: 0.2, ControlNet 0: "Module: canny, Model: control_v11p_sd15_canny [d14c016b], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 100.0, Threshold B: 200.0, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: My prompt is more important", Noise multiplier: 1.05, Version: v1.9.3

With or Without Controlnet?



Original Input Canny Image

With ControlNet



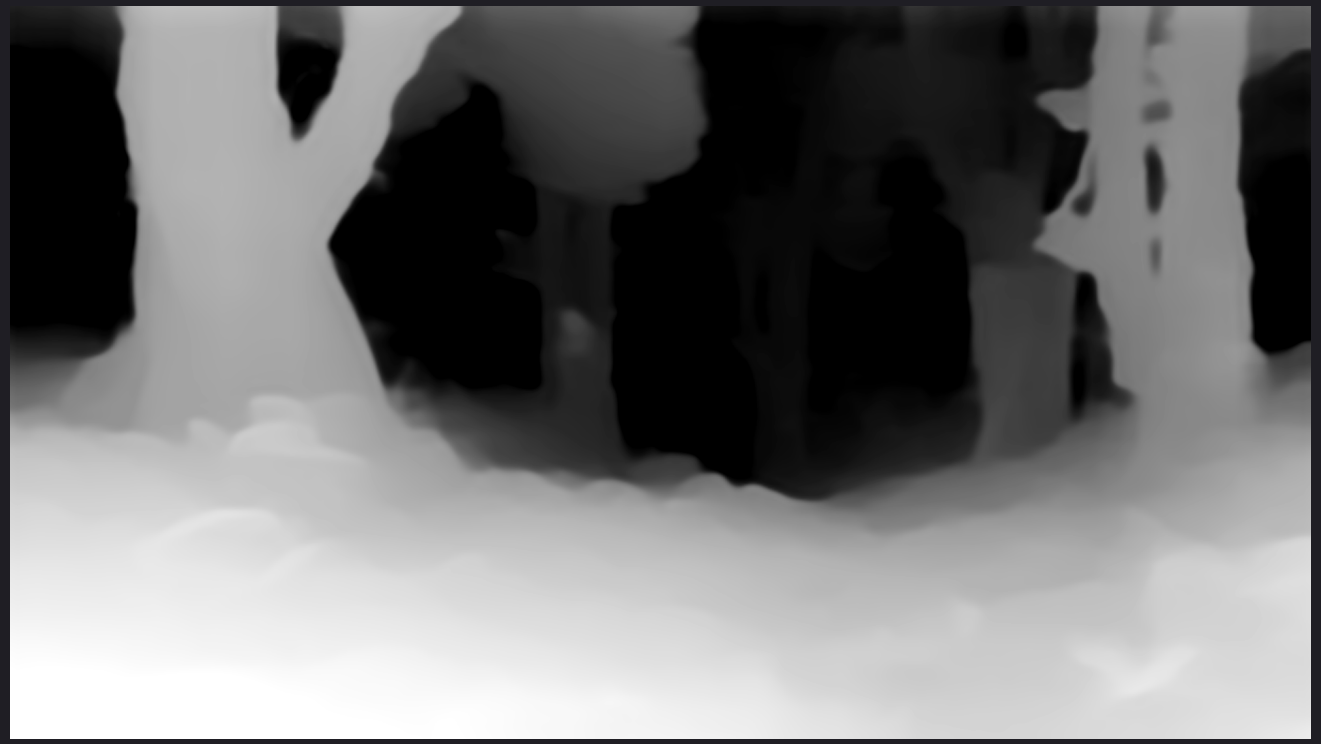
Without ControlNet

A 3D rendering of a Dunhuang mural with deep depth, volumetric lighting, realistic shadows, highly detailed textures, cinematic atmosphere, rendered in Unreal Engine, ultra high definition
Negative prompt: low quality, blurry, deformed, watermark, text

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 7, Seed: 2609786599, Size: 960x473, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Denoising strength: 0.75, Token merging ratio: 0.2, ControlNet 0: "Module: canny, Model: control_v11p_sd15_canny [d14c016b], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 100.0, Threshold B: 200.0, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Noise multiplier: 1.05, Version: v1.9.3

Depth Model

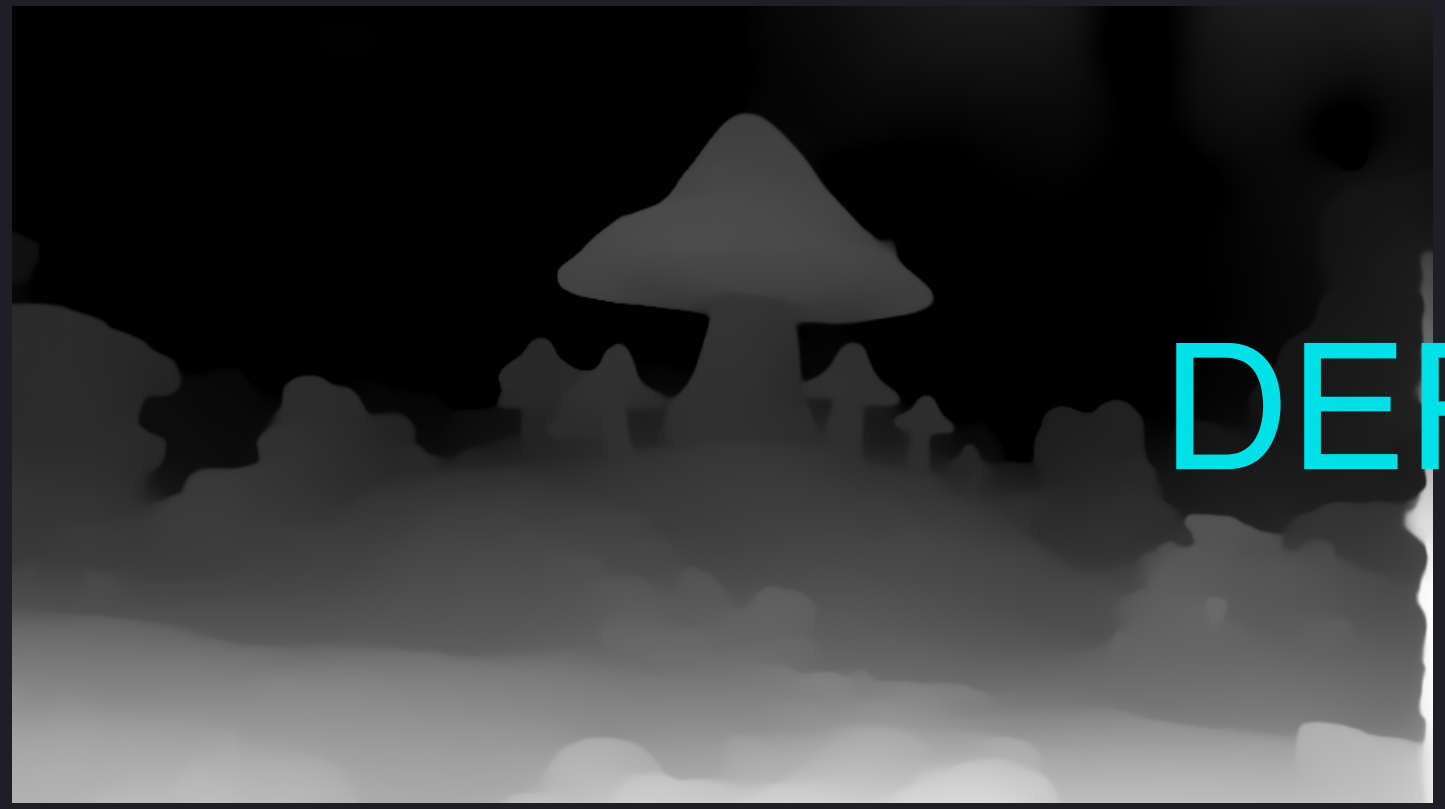
- Function: Adds depth and spatial perception to images.
- Application: By inputting depth maps, we can transform flat images into images with three-dimensional effects, enhancing the dimensionality of the scene.



Depth Model

ORINGINAL IMAGE

Depth Model



DEPTH IMAGE

Depth Model Output

A mystical fairytale forest filled with glowing mushrooms and floating fireflies, tall ancient trees with twisted branches, mist rolling over the forest floor, soft ethereal lighting, enchanted atmosphere, ultra-detailed, high-resolution, fantasy art, volumetric lighting, 8K, Unreal Engine render

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 12, Seed: 2759343917, Size: 1080x696, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Token merging ratio: 0.2, ControlNet 0: "Module: depth_midas, Model: control_v11f1p_sd15_depth [cfd03158], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Version: v1.9.3

Depth Model Output

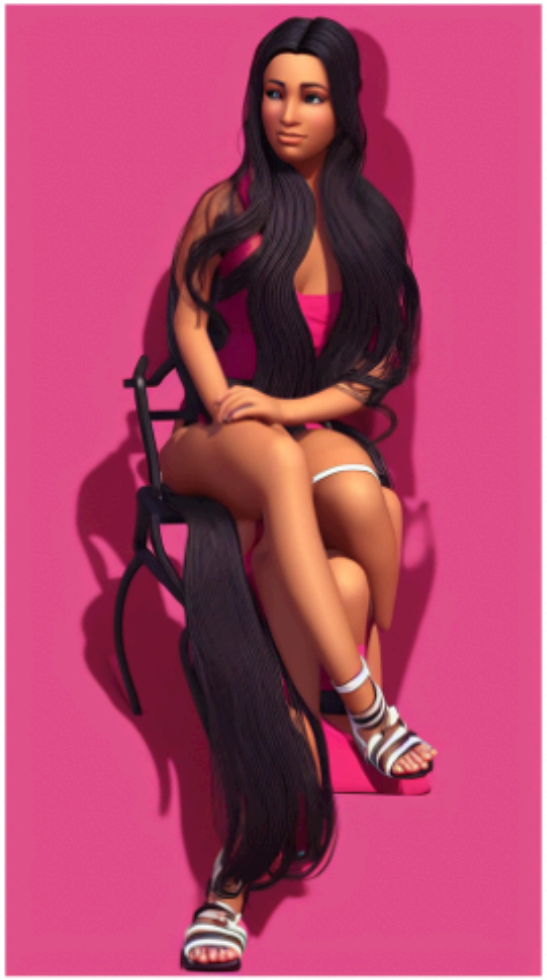
A mystical fairytale forest filled with glowing mushrooms and floating fireflies, tall ancient trees with twisted branches, mist rolling over the forest floor, soft ethereal lighting, enchanted atmosphere, vibrant and dreamy rainbow colors, colorful flora and plants, magical light effects, ultra-detailed, high-resolution, fantasy art, volumetric lighting, 8K, Unreal Engine render
Negative prompt: low quality, blurry, deformed, disfigured, noisy, overexposed, underexposed, dull colors, mediocre color grading, muted tones, flat lighting, grayscale, monochrome, dark atmosphere, watermark, text, people, animals

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 12, Seed: 1265101828, Size: 1080x696, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Token merging ratio: 0.2, ControlNet 0: "Module: depth_midas, Model: control_v11f1p_sd15_depth [cfd03158], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Version: v1.9.3

Comparison between Canny model and Depth model

Original image

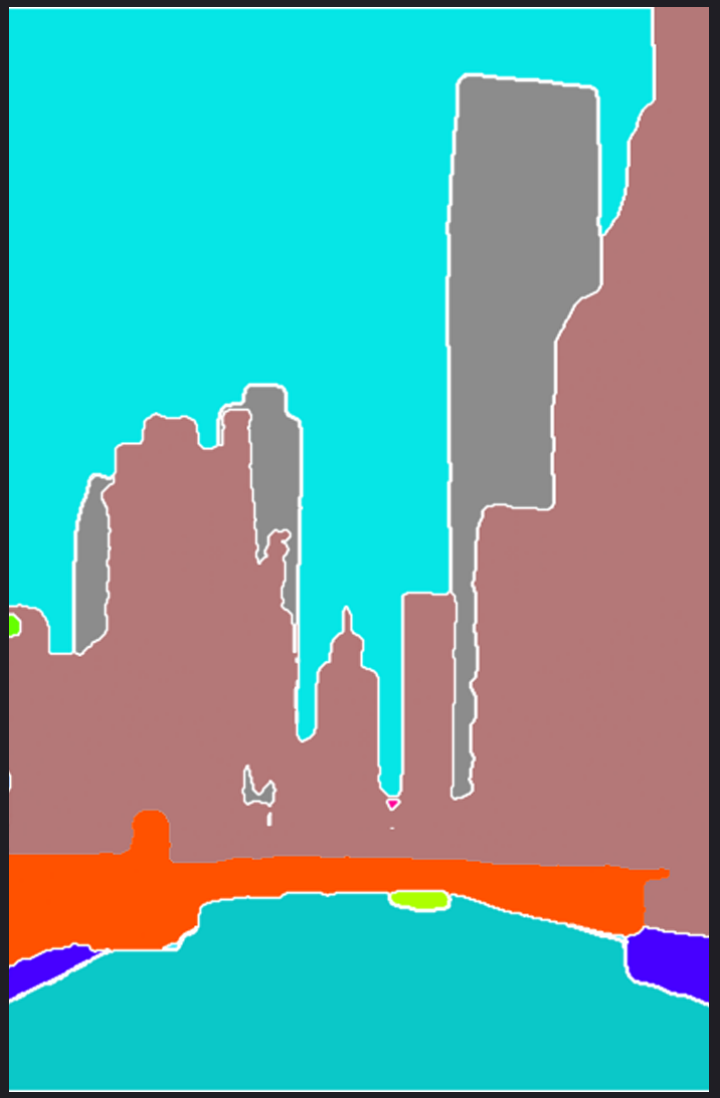
Canny model



Depth model

Segmentation Model

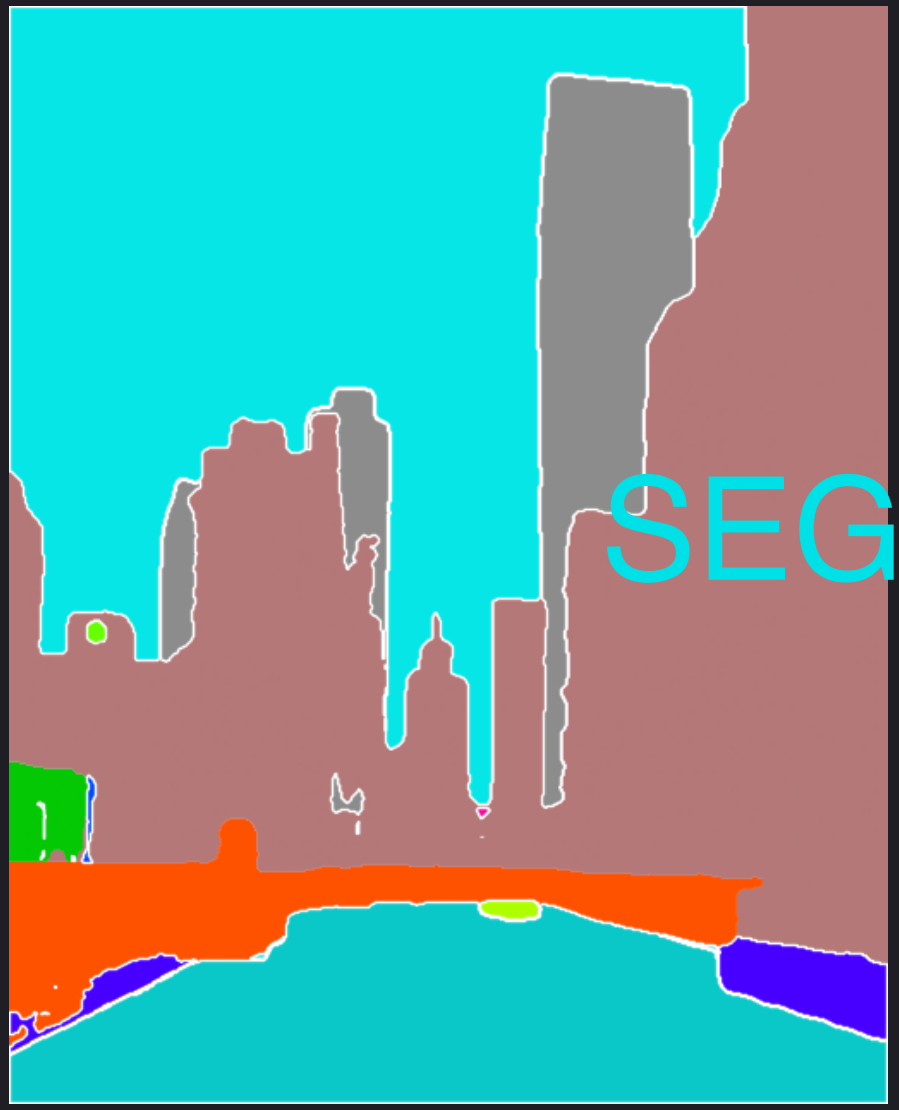
- Function: Performs semantic segmentation, allowing different content or styles to be applied to different regions.
- Application: Enables replacing or applying different artistic styles to regions like the sky, ground, buildings within an image, achieving scene replacement and style mixing.



Segmentation Model

ORINGINAL IMAGE

Segmentation Model



SEGMENTATION IMAGE

Segmentation Model Output

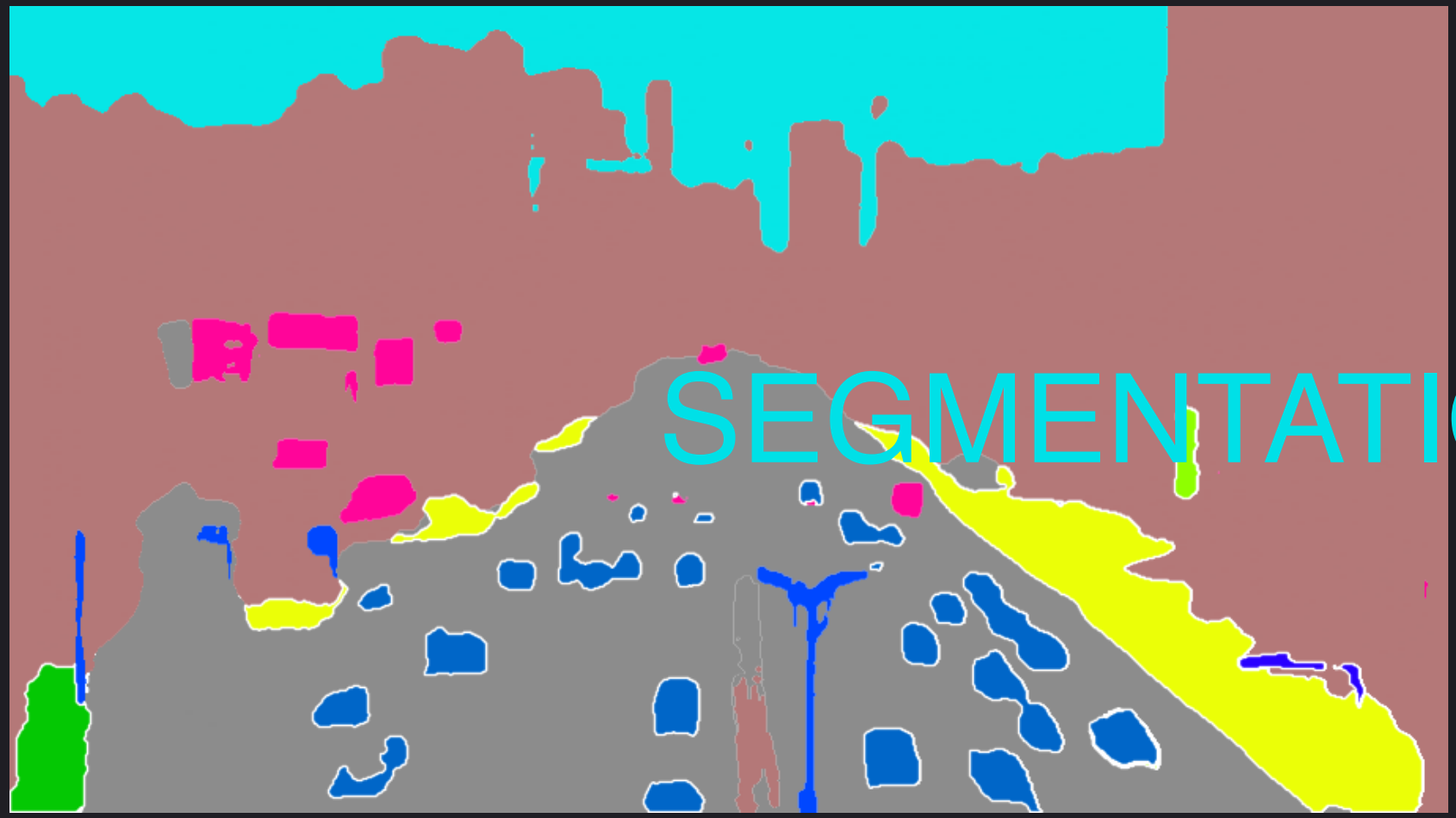
A breathtaking fantasy landscape with a vibrant sunset sky filled with swirling nebula clouds and sparkling stars, the ground covered in lush, luminescent flora with hues of purple and blue, towering crystal formations replacing buildings, majestic floating islands in the background, rendered in a dreamy and colorful 3D style, ultra-detailed, high resolution, soft lighting, cinematic atmosphere, Octane Render, 8K
Negative prompt: low quality, blurry, dull colors, flat lighting, noise, artifacts, watermark, text, overexposed, underexposed

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 12, Seed: 1265101828, Size: 696x1080, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Token merging ratio: 0.2, ControlNet 0: "Module: seg_ofade20k, Model: control_v11p_sd15_seg [e1f51eb9], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Version: v1.9.3

Segmentation Model

ORINGINAL IMAGE

Segmentation Model

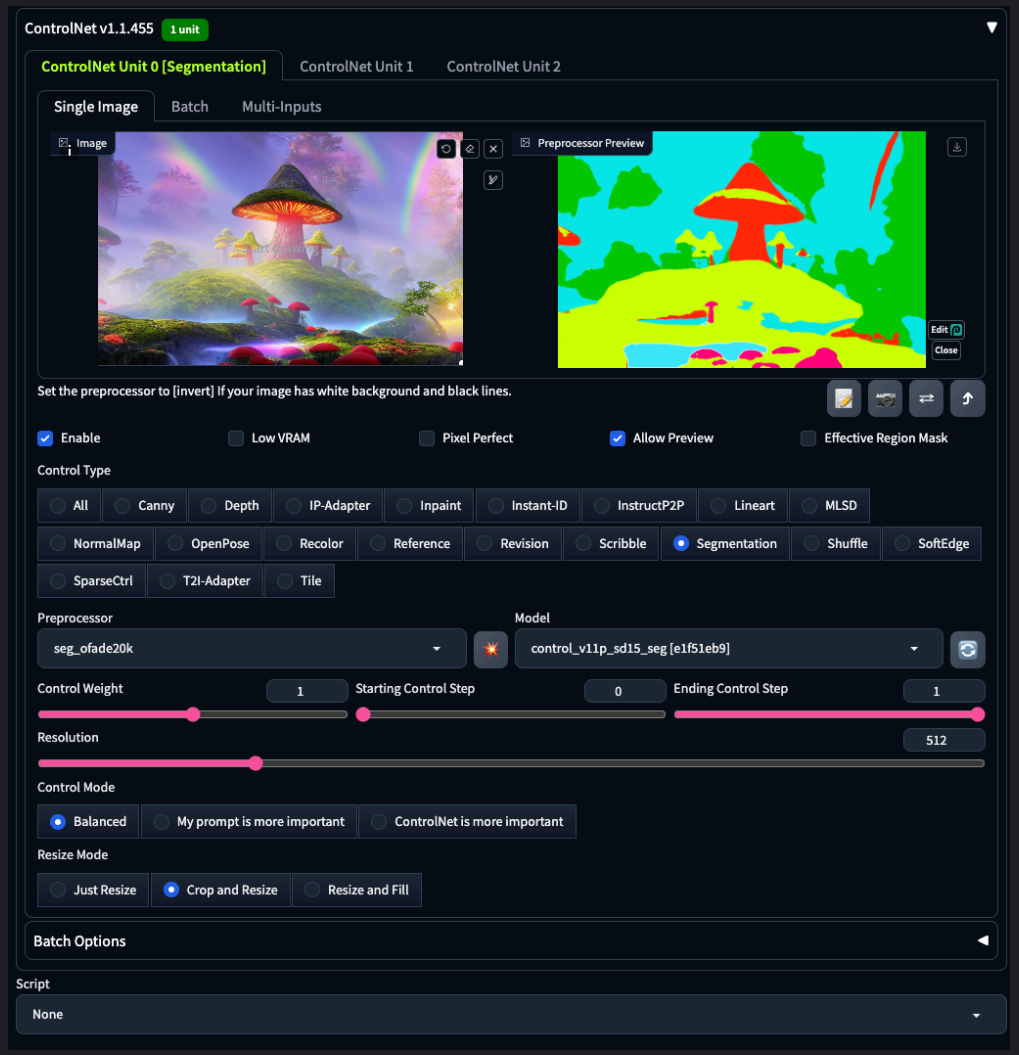


Segmentation Model Output

A breathtaking fantasy landscape with a vibrant sunset sky filled with swirling nebula clouds and sparkling stars, a shimmering river with crystal-clear water flowing where the road used to be, the river reflects the colorful sky above, the ground covered in lush, luminescent flora, towering crystal formations replacing buildings, majestic floating islands in the background, rendered in a dreamy and rainbow color3D style, ultra-detailed, high resolution, soft lighting, cinematic atmosphere, volumetric lighting, Octane Render, 8K
Negative prompt: roads, vehicles, low quality, blurry, dull colors, flat lighting, noise, artifacts, watermark, text, overexposed, underexposed

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 12, Seed: 1265101828, Size: 1080x696, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Token merging ratio: 0.2, ControlNet 0: "Module: seg_ofade20k, Model: control_v11p_sd15_seg [e1f51eb9], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Version: v1.9.3

Segmentation Model



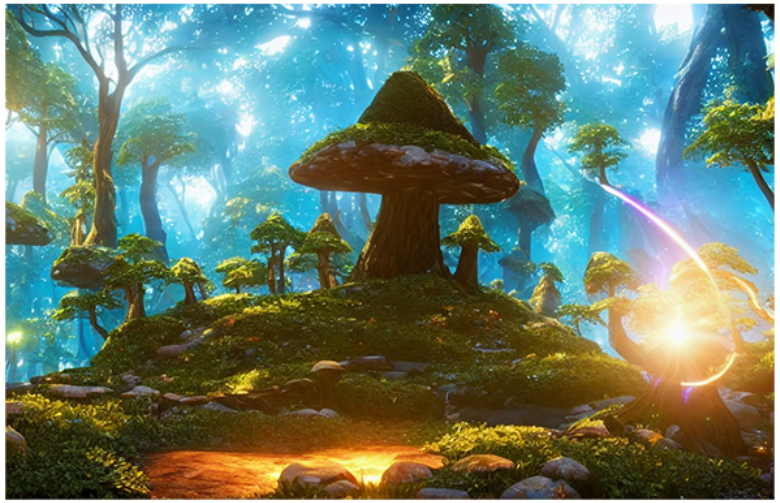
Workflow UI

Output

Comparison between Depth model and Segmentation

Original img

Depth model



Segmentation model

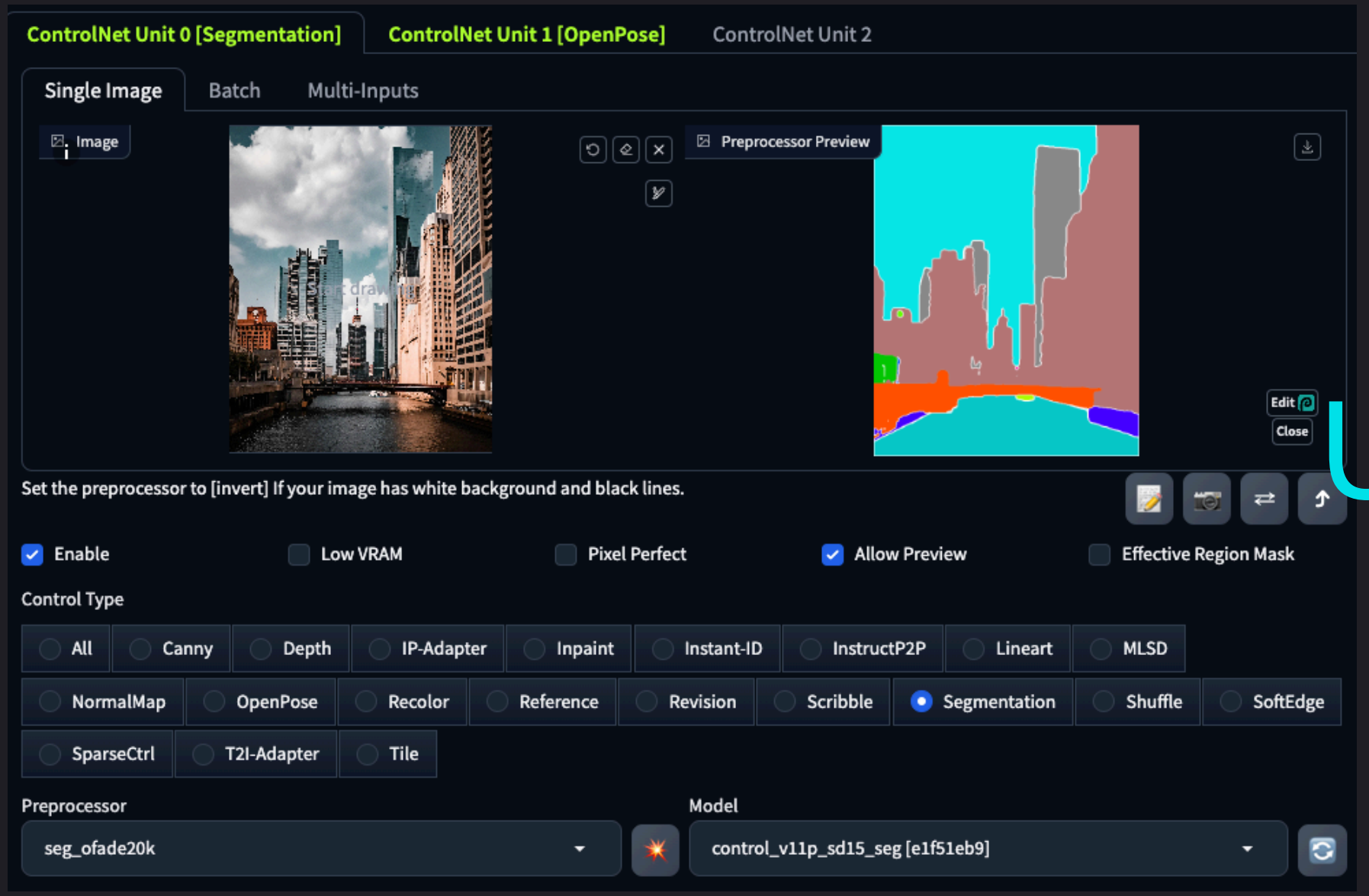
MultiControlNet

- Multiple Inputs: Combine different types of control inputs (e.g., pose, depth, segmentation) for enhanced image manipulation.
- Enhanced Flexibility: Allows for more complex and nuanced control over various elements within an image.
- Improved Precision: Achieve higher accuracy by leveraging the strengths of multiple ControlNet models concurrently.

MultiControlNet Applications

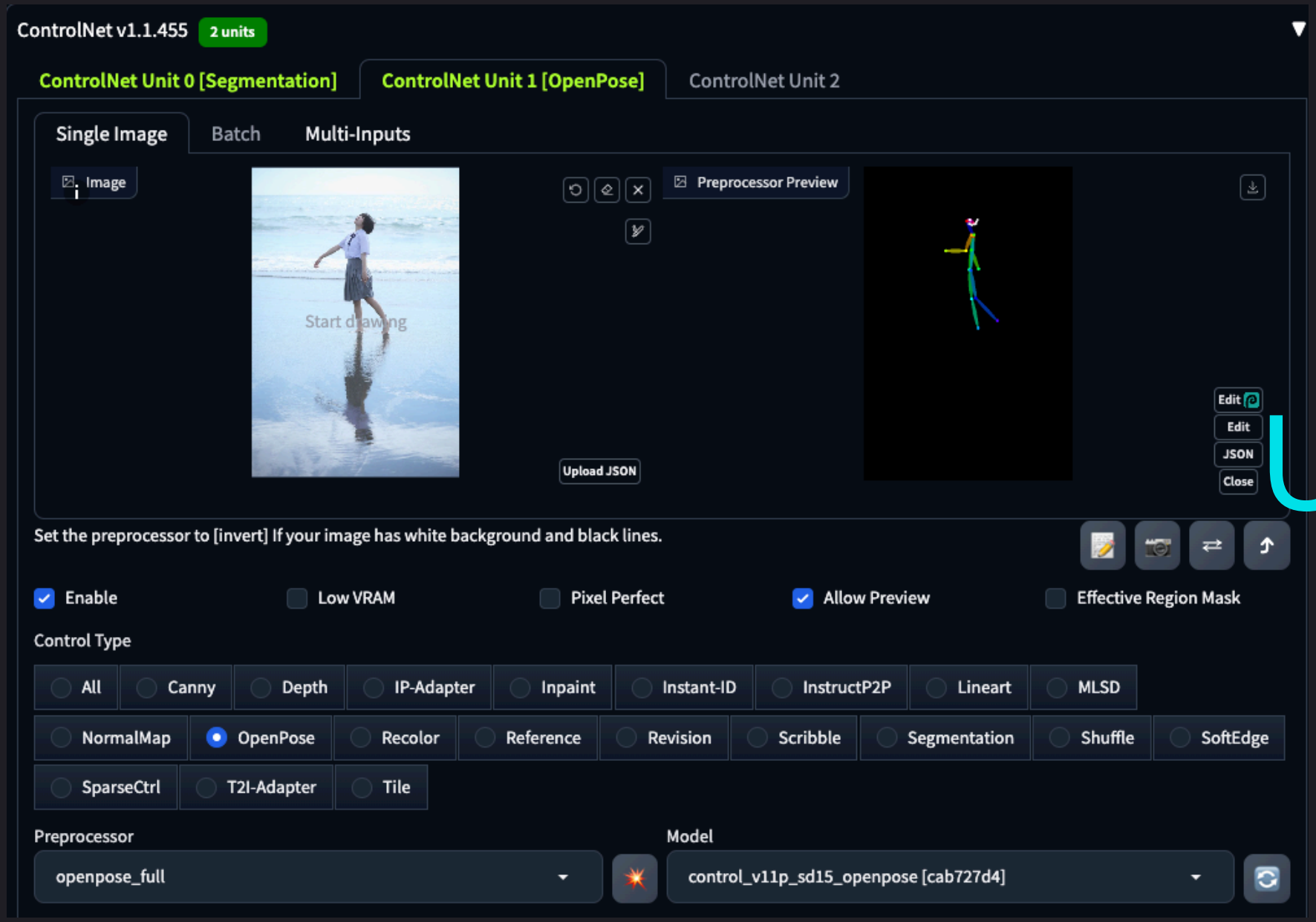
- Advanced Character Design: Control poses, depth, and segmentation simultaneously for detailed character illustrations.
- Scene Composition: Manage multiple elements like foreground, background, and object placement with precision.
- Enhanced Visual Storytelling: Create rich, layered visuals that convey more complex narratives.

MultiControlNet



Unit 0

MultiControlNet



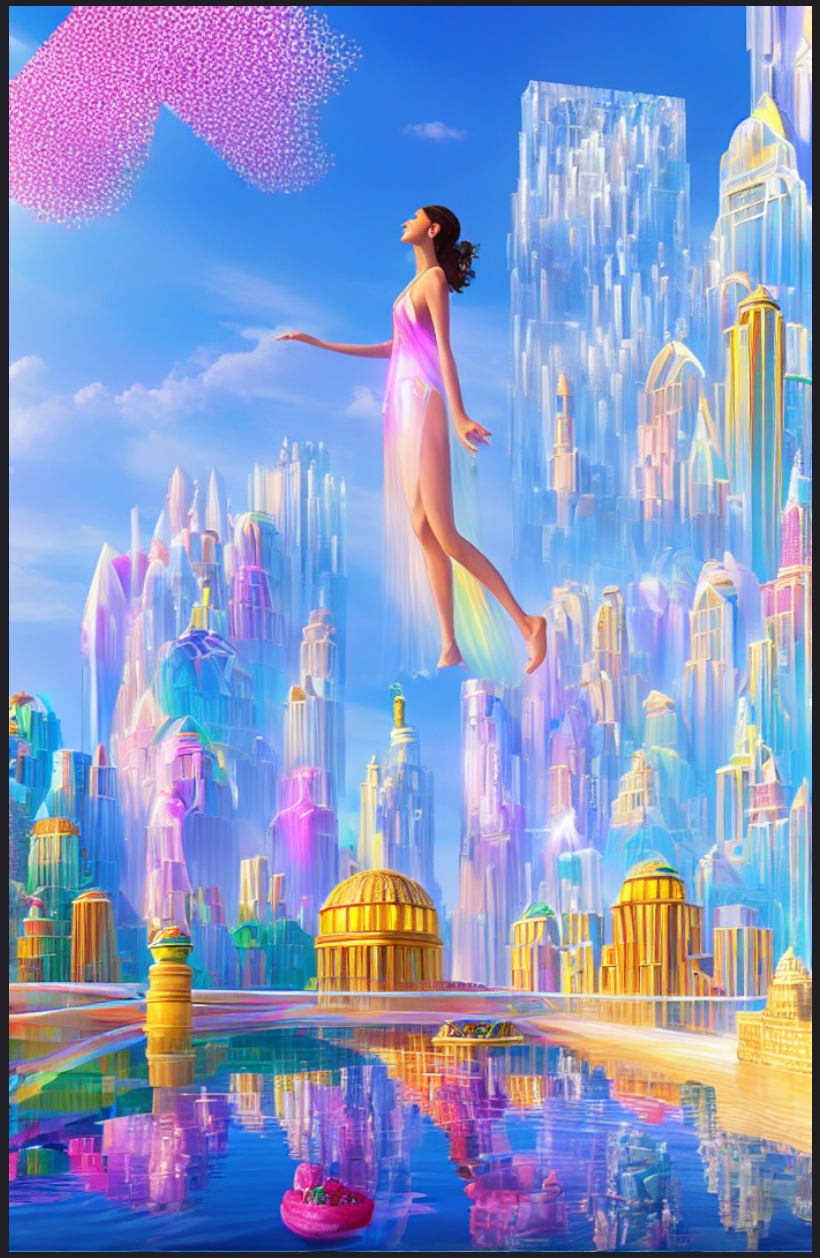
Unit 1

MultiControlNet Output

An enchanting fantasy realm where skyscrapers are transformed into gleaming crystal structures reflecting all the colors of the rainbow, magnificent rainbows stretch across a sky ablaze with vibrant hues, roads have become tranquil, shimmering water surfaces mirroring the kaleidoscopic sky, traffic replaced by ethereal floating plankton emitting a soft, multicolored glow, a girl standing gracefully on a giant, luminous lotus floating atop the iridescent water, surrounded by gentle ripples illuminated by prismatic light, warm and colorful ethereal lighting bathes the scene, enchanted atmosphere, vibrant and dreamy rainbow colors throughout, magical and radiant light effects, ultra-detailed, high-resolution, fantasy art, volumetric lighting, 8K, Unreal Engine render

Negative prompt: low quality, blurry, deformed, disfigured, noisy, overexposed, underexposed, muted colors, dull colors, grayscale, monochrome, dark atmosphere, watermark, text, traffic, animals

Steps: 20, Sampler: DPM++ 2M, Schedule type: Karras, CFG scale: 15, Seed: 1519522183, Size: 696x1080, Model hash: cc6cb27103, Model: v1-5-pruned-emaonly, Token merging ratio: 0.2, ControlNet 0: "Module: seg_ofade20k, Model: control_v11p_sd15_seg [e1f51eb9], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", ControlNet 1: "Module: openpose_full, Model: control_v11p_sd15_openpose [cab727d4], Weight: 1.0, Resize Mode: Crop and Resize, Processor Res: 512, Threshold A: 0.5, Threshold B: 0.5, Guidance Start: 0.0, Guidance End: 1.0, Pixel Perfect: False, Control Mode: Balanced", Version: v1.9.3



MAT 255 Fall 2024
Final Project

Shaw Xiao