Figure 9-17. Outpainting in Img2Img

As you can see in Figure 9-16, with the extra castle being added to the sky, the potential for hallucination is high and the quality can be low. It often takes a lot of experimentation and iteration to get this process right. This is a similar technique to how early adopters of generative AI would add extra empty space on the sides of photos in photoshop, before inpainting them to match the rest of the image in Stable Diffusion. This technique is essentially just inpainting with extra steps, so all of the same advice listed above applies. This can be quicker than using the outpainting functionality in AUTOMATIC1111 because of the poor quality and limitations of not having a proper canvas.

ControlNet

Using prompting and Img2Img or base images it's possible to control the style of an image, but often the pose of people in the image, composition of the scene, or structure of the objects will differ greatly in the final image. ControlNet is an advanced way of conditioning input images for image generation models like Stable Diffusion. It allows you to gain more control over the final image generated through various techniques like edge detection, pose, depth, and many more. You upload an image you want to emulate, and use one of the pre-trained model options for processing the image to input alongside your prompt, resulting in a matching image composition with a different style (Figure 9-18, from the ControlNet paper).

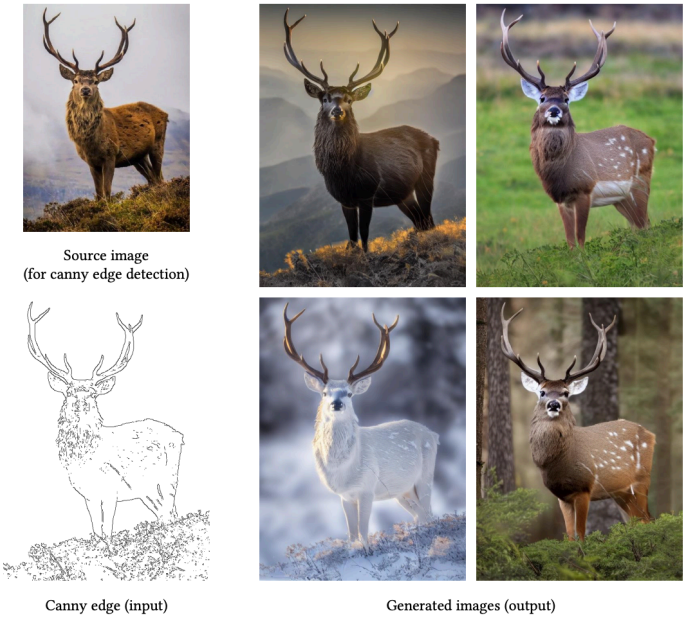

Figure 9-18. ControlNet Stable Diffusion with Canny edge map

What's referred to as ControlNet is really a series of open-source models released following the paper "Adding Conditional Control to Text-to-Image Diffusion Models" (Zhang and Agrawala, 2023). While it is possible to code this in Python and build your own user interface for it, the quickest and easiest way to to get up and running is via the ControlNet extension for AUTOMATIC1111. As of the time of writing, not all ControlNet methods are available for SDXL, so we are using Stable Diffusion v1.5 (make sure you use a ControlNet model that matches the version of Stable Diffusion you're using). You can install the extension following these instructions:

1. Navigate to the Extensions tab and click the sub tab labelled Available
2. Click the *Load from* button.
3. In the Search box type `sd-webui-controlnet` to find the Extension.
4. Click Install in the Action column to the far right.
5. Web UI will now download the necessary files and install ControlNet on your local version of Stable Diffusion.

If you have trouble executing the preceeding steps you can try the following alternate method:

1. Navigate to the Extensions tab and click Install from URL sub tab.
2. In the URL field for the Git repository, paste the link to the extension: *https://github.com/Mikubill/sd-webui-controlnet*
3. Click Install.
4. WebUI will download and install the necessary files for ControlNet.

Now that you have ControlNet installed, restart AUTOMATIC1111 from your terminal or command line, or visit Settings and click "Apply and restart UI".

The extension will appear below the normal parameter options you get for Stable Diffusion, in an accordion tab (Figure 9-18). You first upload an image, then click enable before selecting the ControlNet preprocessor and

model you wish to use. If your system has less than 6 GB of VRAM (Video Random Access Memory), you should check the Low VRAM box. Depending on the task at hand you might want to experiment with a number of models, and make adjustments to the parameters of those models, in order to see which gets results.

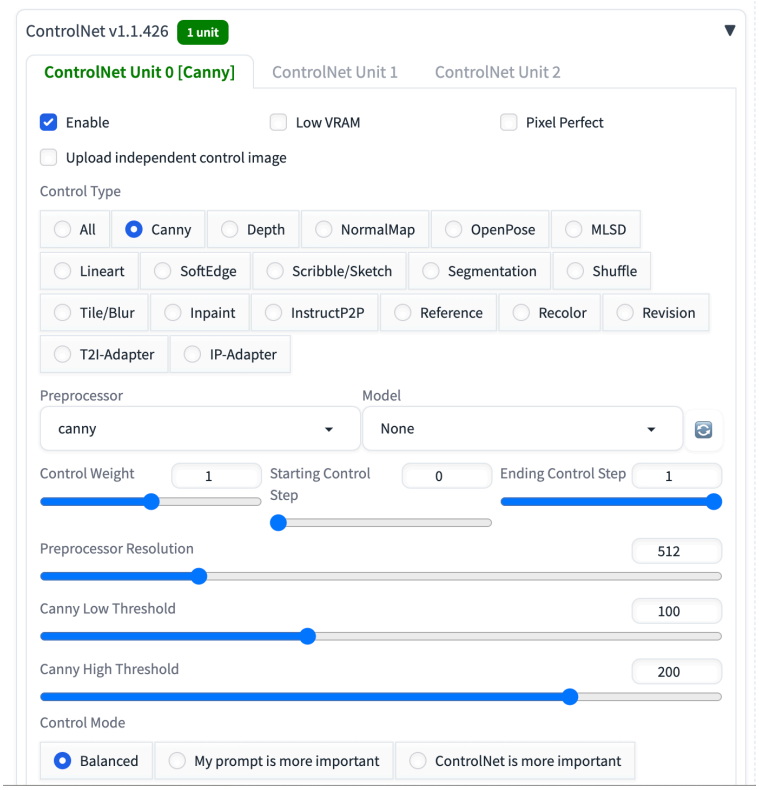

Figure 9-19. ControlNet extension interface in AUTOMATIC1111

Control Weight is analogous to prompt weight or influence, similar to putting words in brackets with a weighting `(prompt words: 1.2)`, but for the ControlNet input. The Starting Control Steps and Ending Control Steps are when in the diffusion process the ControlNet applies, by default from start to finish (0 to 1), akin to prompt editing / shifting such as `[prompt words::0.8]` (apply this part of the prompt from the beginning until 80% of the total steps are complete). Because the image diffuses from larger elements down to finer details, you can achieve different results by controlling where in that process the ControlNet applies; for example; removing the last 20% of steps (Ending Control Step = 0.8) may allow the model more creativity when filling in finer detail. The Preprocessor Resolution also helps maintain control here, determining how much fine detail there is in the intermediate image processing step. Some models have their own unique parameters, such as the Canny Low and High Thresholds, which determine what pixels constitute an *edge*. Finally the Control Mode determines how much the model follows the ControlNet input relative to your prompt.

When you first install ControlNet you won't have any models downloaded, so for them to populate in the dropdown you should install them by downloading them from the models page, and then dropping them in the *Models* > *ControlNet* folder. If you're unsure of which model to try, start with Canny edge detection as it is the most generally useful. Each model is relatively large (in the order of a few gigabytes) so only download the ones you plan to use. Below are examples from some of the more common models. All images in this section are generated with the `DPM++`

`SDE Karras` sampler, a CFG scale of 1.5, Control Mode set to Balanced, Resize Mode set to Crop and Resize (the uploaded image is cropped to match the dimensions of the generated image, 512 x 512), and 30 sampling steps, with the default settings for each ControlNet model. Version 1.5 of Stable Diffusion was used as not all of these ControlNet models are available for Stable Diffusion XL at the time of writing, but the techniques should be transferrable between models.

Canny edge detection creates simple, sharp pixel outlines around areas of high contract. It can be very detailed and give excellent results, but can also pick up unwanted noise and give too much control of the image to ControlNet. In images where there is a high degree of detail that needs to be transferred to a new image with a different style, Canny excels and should be used as the default option. For example, redrawing a city skyline in a specific style works very well with the Canny model, as we did with an image of New York City (by Robert Bye on Unsplash) in Figure 8-17:

Input:

```
new york city by studio ghibli
```

The output is shown in Figure 9-20.

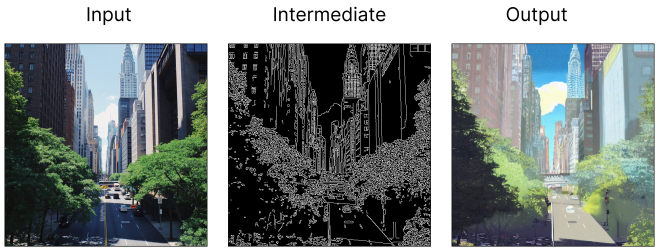

Figure 9-20. ControlNet Canny

Sometimes in traditional `img2img` prompting, some elements of an image get confused or merged, because Stable Diffusion doesn't understand the depth of those objects in relation to each other. The Depth model creates a depth map estimation based on the image, which provides control over the composition and spatial position of image elements. If you're not familiar with depth maps, whiter areas are closer to the viewer, and blacker are further away. This can be seen in Figure 8-18, where an image of a band (by Hans Vivek on Unsplash) is turned into an image of soldiers with the same positions and depth of field:

Input:

```
us military unit on patrol in afghanistan
```

The output is shown in Figure 9-21.

Figure 9-21. ControlNet Depth

The Normal model creates a mapping estimation that functions as a 3D model of objects in the image. The colors red, green, and blue are used by 3D programs to determine how smooth or bumpy an object is, with each color corresponding to a direction (left/right, up/down, close/far). This is just an estimation, however, so it can have unintended consequences in some cases. This method tends to excel if you need more textures and lighting to be taken into consideration, but can sometimes offer too much detail in the case of faces, constraining the creativity of the output. In Figure 9-21 a woman playing a keyboard (by Soundtrap on Unsplash) is transported back in time to the Great Gatsby era:

Input:

```
woman playing piano at a great gatsby flapper party, 1920s,
symmetrical face
```

The output is shown in Figure 9-22.

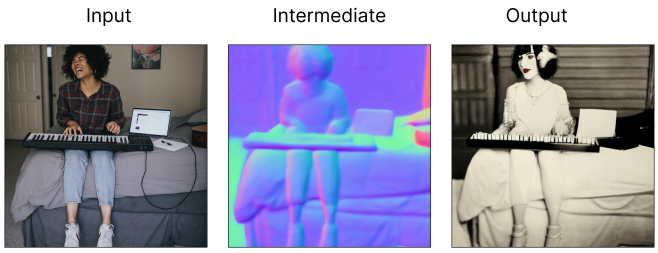

Figure 9-22. ControlNet Normal

The OpenPose method creates a skeleton for a figure by determining its posture, hand placement and facial expression. For this model to work you typically need to have a human subject with the full body visible, though there are portrait options. It is very common practice to use multiple OpenPose skeletons and compose them together into a single image, if multiple people are required in the scene. Figure 9-23 transposes the Mona Lisa's pose onto an image of Rachel Weisz:

Input:

```
painting of Rachel Weisz
```

The output is shown in Figure 9-23.

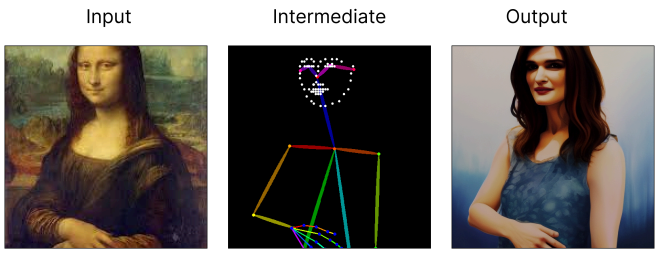

Figure 9-23. ControlNet OpenPose

The MLSD (Mobile Line Segment Detection) technique is quite often used in architecture and interior design, as it's well suited to tracing straight lines. Straight lines tend only to appear in man-made objects, so it isn't well suited to nature scenes (though it might create an interesting effect). Man made objects like houses are well suited to this approach, as seen in the image of a modern apartment (by Collov Home Design on Unsplash) reimagined for the *Mad Men* era, in Figure 9-24:

Input:

```
1960s mad men style apartment
```

The output is shown in Figure 9-24.

Figure 9-24. ControlNet MLSD

The SoftEdge technique, also known as HED (Holistically-nested Edge Detection), is an alternative to Canny edge detection, creating smoother outlines around objects. It is very commonly used and provides good detail like Canny, but can be less noisy and deliver more aesthetically pleasing results. This method is great for stylizing and recoloring images, and it tends to allow for better manipulation of faces compared to Canny. Thanks to ControlNet, you don't need to enter too much of a detailed prompt of the overall image, and can just prompt for the change you want to see. Figure 9-25 shows a reimagining of Vermeer's Girl with a Pearl Earring, with Scarlett Johansson:

Input:

```
scarlett johansson, best quality, extremely detailed
```

Negative:

```
monochrome, lowres, bad anatomy, worst quality, low quality
```