Chapter 8. Standard Practices for Image Generation with Midjourney

In this chapter you'll use standardized techniques to maximize the output and formats from diffusion models. You'll start by tailoring the prompts to explore all of the common practices used for image generation. All images are generated by Midjourney v5, unless otherwise noted. The techniques discussed were devised to be transferrable to any future or alternative model.

Format Modifiers

The most basic practice in image generation is to specify the format of the image. AI image models are capable of deploying a wide variety of formats, from stock photos, to oil paintings, or even ancient Egpytian hieroglyphics. The image often looks completely different depending on the format, including the style of the objects or people generated in the image. Many of the images in the training data are stock photos, and this is

also one of the most commercially important image categories for image generation.

Input:

```
a stock photo of a business meeting
```

The output is shown in .

Figure 8-1. Stock photo of a business meeting

The ability to generate infinite royalty-free stock photos for free with open-source models like Stable Diffusion, or for a very low cost with DALL-E or Midjourney, is itself a game-changer. Each of these images is unique (though may contain similarities to existing images) and therefore they look more premium than reusing the same free stock photos available to everyone else. However, you no longer need to be limited to the

stock photography format. If your blog post or website imagery would look better with something more artistic, you can do that with essentially no limits:

Input:

```
an oil painting of a business meeting
```

The output is shown in Figure 8-2.

Figure 8-2. Oil painting of a business meeting

SPECIFY FORMAT

The format we specify significantly modifies the results we get from our AI model. Specifying format also improves the reliability of our prompts in terms of giving us the type of visual we require.

There's no real limit to how far you want to take this technique, and this is one of those domains where it would have helped to go to art school. If you know the name of a specific technique or detail you want to see in your image – for example Impasto, a technique used in oil painting, where paint is laid on an area of the surface thickly, leaving visible brush strokes – you can reference it in the prompt to get closer to your desired result. Google maintains a comprehensive list of popular artists and art movements that many find useful.

Input:

```
an oil painting of a business meeting, textured oil-on-canvas
using thick impasto and swirling dynamic brushstrokes
```

The output is shown in Figure 8-3.

Figure 8-3. Oil painting of a business meeting with Impasto

The oil painting of the business meeting is now far more visually interesting and potentially more appealing, depending on your audience. Traditionally, one of the reasons businesses migrated to using stock photography is that it was cheaper than commissioning a painting, but that limitation no longer applies with AI. We can generate essentially any format we like, for example an ancient Egyptian hieroglyph of a business meeting:

Input:

```
an ancient egyptian hieroglyph of a business meeting
```

The output is shown in Figure 8-4.

Figure 8-4. Ancient Egyptian hieroglyph of a business meeting

The thing to watch out for with modifying the format is that the style of the image, and even the contents, tend to match what was associated with that format in the training data. For example, in our oil painting there aren't any computers, because they don't often appear in oil painting. Similarly in our hieroglyph the participants in the meeting are wearing ancient Egyptian headresses. Often you'll need to combine format modifiers with the other proceeding techniques in order to arrive at what you want.

Art Style Modifiers

One of the great powers of AI image models is their ability to replicate any popular art style or artist. The most common examples shared on social media and AI demos are images in the style of Van Gogh, Dali, or Picasso, as well as the art movements they were a part of, respectively Post-Impressionism, Surrealism, and Cubism. AI art communities have also become influential in determining what contemporary art styles become popular, as is the case with a polish digital artist called Greg Rutkowski, known for his fantasy style. However, many artists have taken a stand against AI art, and there is a legal gray area around whether imitating a living artist's style is considered *fair use* under copyright law. We recommend AI artists exercise caution when generating AI art in the distinctive style of any living artist, and instead stick to artists who died 100 years ago as a rule of thumb (seek legal counsel for any planned commercial use):

Input:

```
illustration of a dragon, in the style of Alice's Adventures in Wonderland
by Lewis Carroll
```

The output is shown in Figure 8-5.

Figure 8-5. Illustration of a dragon, in the style of Lewis Carroll

GIVE DIRECTION

Evoking an artist's name or the name of an art movement is a shortcut towards delivering a specific visual style. So long as the artist or art movement has enough examples in the training data, their nature can be emulated.

In evoking an artist's style you're effectively shortcutting to a part of the *latent space,* the multi-dimensional universe of potential model outputs, filtering down to your desired style. Traversing to nearby locations from there can help you arrive at a more pleasing destination than you could get to with random trial and error.

Reverse-Engineering Prompts

If you didn't go to art school, or don't know much about film or photography, it can be daunting to try and figure out the art styles, formats, and artists you want to take advantage of. Often you see a picture you like and have no way of describing it in enough detail to recreate it with a prompt. Thankfully Midjourney's `Describe` functionality allows you to reverse-engineer a prompt from an image by typing `/describe` and then uploading the image. It works for both AI generated images and also normal images from other sources too, as shown in Figure 8-6, using one of the stock photos from Chapter 1.

Figure 8-6. Midjourney Describe, Photo by Mimi Thian on Unsplash

Midjourney gives you four options with various artists, art styles, modifiers, and other words, including an estimation of what is happening in the image and what subjects or elements are contained. For example, Midjourney correctly identifies a family posed with their luggage next to a car, in Figure 8-6. You can select the option you want by number, and Midjourney will generate an image with that prompt, in the same style as the original. There is similar open-source technology available named CLIP Interrogator, though the richness of the prompt and ability to replicate the style of the uploaded image is lacking compared to Midjourney.

Quality Boosters

One trick that works for image models is to add words that are associated with quality into the prompt. Some art styles are more aesthetic than others, but there is a set of words, known as *quality boosters*, that seem to improve the image quality without greatly affecting the style, like "4k", "very beautiful", and "trending on artstation". Generative models aren't trying to make high quality images, they're trying to imitate training sets with a wide variety of styles and qualities: if you want high quality images you must explicitly ask for it. Start with the subject of your prompt, for example a space whale, and add a modifier to the end, separated by a comma (as in Figure 8-7):

Input:

```
a space whale, trending on artstation
```

Output:

Figure 8-7. Space whale, trending on artstation

Using quality boosters can help improve the aesthetics of an image through the addition of one or two words to the prompt, without changing the overall style of the image by as much.

The reason these labels work is that they were associated with quality in the training data. When AI image models were trained, they reportedly ingested images from popular design portfolio websites, such as ArtStation, Behance, and DeviantArt. Therefore the model can approximate that an image that was "trending on artstation" was of higher aesthetic value than normal. Note that sometimes style seeps through that may not be aligned with your creative vision. For example, ArtStation contains a lot of digital art of spaceships, and that perhaps explains why the space whale in Figure 38 somewhat resembles a space ship. For a list of quality boosters, art styles, and artists, visit this template created by one of the authors: Prompt Engineering Template. Google also compiles a comprehensive list of art movements which can be useful for educating yourself on the names of styles you find appealing.

Negative Prompts

Often two concepts are so intertwined in the training data that they appear together frequently when generating images of one of the concepts, even if that's not what you specified or intended. For example when you ask for oil paintings you often get the accompanying frame, and surrounding wall, because that's what's in the images for a large number of museum collections of these paintings.

In Midjourney and Stable Diffusion there is the ablity to add *negative prompts*, which allow you to specify what you don't want in the image. Negative prompts can be used to effectively separate two intertwined concepts and ensure your image doesn't contain anything you were hoping to avoid. Taking the example of oil paintings and frames, you can add `--no` to the end of the prompt, and anything in a comma-separated list after that flag will be negated from the prompt. To fix your frames prob-

lem, add "frame" and "wall" as a negative prompt, as is shown in Figure 8-8:

Input:

```
oil painting in the style of rembrandt --no frame, wall
```

The output is shown in Figure 8-8.

Figure 8-8. Oil painting in the style of rembrandt without frames

GIVE DIRECTION

Negative prompts can negate unwanted concepts in images, directing the model away from areas you are trying to avoid. This doesn't always work as intended, as often the concepts are too well correlated, but when it does it can lead to interesting places.

Negative prompting isn't fully reliable, but can be useful in a wide variety of scenarios. One creative use of this technique is to add the name of a celebrity as a negative prompt, in order to decrease the factors most associated with them. The famous actress Karen Gillan has red hair and has a

conventionally feminine look, and therefore can be used to make a subject less likely to have red hair or look conventionally feminine:

Input:

```
a scottish female astronaut --no Karen Gillan
```

The output is shown in Figure 8-9.

Figure 8-9. A less conventionally feminine, less red-haired Scottish female astronaut

You can also get very creative and unpredictable outcomes with this technique by taking two inseparable concepts and seeing what happens when you separate them. For example, try taking your favorite cartoon and removing the cartoon style, as depicted in Figure 8-10 with Homer Simpson.

Input:

```
homer simpson --no cartoon
```

The (horrifying) output is shown in Figure 8-10.

Figure 8-10. Homer Simpson without his trademark cartoon style

One of the more common historical use cases for negative prompts traditionally is to correct some of the issues with disfigured hands, explicit body parts, or odd-looking eyes, all problems early AI models suffered from. Prompt engineers would add words like `nsfw, elongated body, too many digits, not enough fingers, teeth` to the negative prompt in an attempt (often in vain) to guide the model away from these spaces. While still necessary for older or lesser models like Stable Diffusion v1.5, from Midjourney v5 and Stable Diffusion XL onwards this is mostly a solved problem. State-of-the-art models are now more than ca-

pable of developing normal-looking images of hands, eyes, and bodies without relying on negative prompts.

Weighted Terms

Negative prompts are useful if you want to completely negate something, but often you just want to dial it down. In order to mix and match different concepts it can be helpful to have control over how much of each you want.

By default all words in a prompt have an equal weighting of 1, although words at the beginning of the prompt have a greater effect, which is why we typically put the subject of our image there by convention i.e. `painting of the golden gate bridge`. You can change the weights of sections of the prompt in Midjourney by adding a *hard break* with two colon characters `::` and a number denoting the new weight. With this method you can make an image that is primarily Van Gogh but with a dash of Dali:

Input:

```
painting of the golden gate bridge::1 in the style of van
gogh::0.8, in the style of dali::0.2
```

The output is shown in Figure 8-11.

To see how weights affect the resulting image, you can conduct a grid search by systematically testing each combination of weights at a specific granularity. In this example the weights changed in 0.2 increments between the two artists from 0 to 1.

The output is shown in Figure 8-12.

Figure 8-12. Permutations grid of weights

Weights introduce many possible combinations in a prompt, which can be time consuming to iterate through one at a time. A grid search approach of systematically generating many possible combinations is recommended to identify where the ideal mix of weightings aligns with your preferences.

Weights can go higher than 1 as needed for emphasis, or lower if you want to de-emphasize something. You can also add negative weights to the prompt to remove that aspect to varying degress. The `--no` parameter used for negative prompts is actually just a shortcut for adding `::-0.5` to that section of the prompt. If you are struggling with something appearing in the image that you don't want, try stronger negative weights instead of negative prompts. Using the prior example you could strip any Van Gogh influence out of Dali's work by adding a -1 weight to Van Gogh, and dialing up the Dali weight to 5:

Input:

```
painting of the golden gate bridge::1 in the style of van
gogh::-1, in the style of dali::5
```

The output is shown in Figure 8-13.

Figure 8-13. Painting of the golden gate bridge

Weights can be a powerful tool in remixing different styles or emphasizing specific elements. There are many permutations of weights and therefore a more systematic approach must be taken to find an aesthetically interesting space to play in.

Prompting with an Image

Many AI image generation tools let you prompt the model not just with text, but with an image. Supplying an example image of what you're going for can give you a great baseline for building something more unique and original, while still matching the style you need. In the Stable Diffusion community this is called `Img2Img`*, whereas in Midjourney you simply link to an image in the prompt. The best way to do this is to upload the image into Discord first (for example, this photo by Jessica Hearn on Unsplash), then right click and select Copy Link, as depicted in Figure . You can later paste that link into the prompt to use as the base image*

.Copying the link from an image uploaded to Discord

The copied link then should be pasted at the beginning of the Midjourney prompt, and accompanied by your text prompt. You don't need to be as

descriptive now that you have given a base image to go off (a picture is worth 1,000 words). The image won't match exactly, but will be similar to the point of being recognizeable if you know what image was supplied and how it was modified by the prompt:

Input:

```
https://s.mj.run/XkIHsYIdUxc in the style of the great gatsby
```

The output is shown in Figure 8-14.

Figure 8-14. Stock photo in the style of The Great Gatsby

This technique works wherever you want a similar vibe, scene, or composition to an image you know of. You can also blend multiple images together with `/blend` to get something quite unique, and even use the resulting image as input for another prompt. For convenience, there is the `--iw` parameter, which acts the same as separating the image from the rest of the prompt with `::` and setting the weight. Prompting with an image is also a common technique for AI video generation with tools such as RunwayML and Pika Labs, given the general unreliability of text to video generation, and because it gives you an opportunity to iterate on the style of the scene without waiting for a whole video to generate and render.

PROVIDE EXAMPLES

The quickest and easiest way to get the image you desire is to upload an image that you want to emulate. This is similar in concept to a one-shot prompt in the text generation space, and is similarly useful in guiding the model towards the right output.

Inpainting

Working with AI image generation tools is always an iterative process. Rarely do you get the complete final image on the first try. There are usually artifacts that you want to address, or styles that you want to change. For example, say you had generated an image with Midjourney of a woman in a 1920s style flapper dress, but wanted to change what she was wearing without regenerating the entire image.

The solution is *inpainting*, which is available in most implementations of Stable Diffusion, in Midjourney via a feature called *Vary region*, and with Adobe Photoshop's *Generative fill*. However, DALL-E pioneered this functionality, and it is still the author's personal preference in terms of the quality of the results. To demonstrate this functionality, first you upload

an image to DALL-E in OpenAI Labs, then erase the part of the image you want to re-generate.

Figure 8-15 shows an image generated by Midjourney with the accompanying prompt below. It is loaded into DALL-E's inpainting canvas (currently using DALL-E 2) and has had the dress part of the image erased using the inpainting brush, ready for inpainting:

Input:

```
photograph of glamorous woman in a 1920s flapper party,
wearing a sequin dress, wide angle, in color, 3 5 mm, dslr
```

The uploaded image with parts erased is shown in Figure 8-15.

Then you add a prompt for what you want to generate within that space.
The common advice is to prompt for what you want the whole image to
be, but in our experience narrowing down the prompt to just what you
want in the erased part of the image gets better results, as in Figure 8-16
which focuses on the dress itself.

Input:

```
van gogh style dress, inspired by starry night, blue and
yellow swirls, extremely detailed, very well lit, studio
light, 3.5 mm, dslr
```

The output is shown in Figure 8-16.

Figure 8-16. Van Gogh style dress

DIVIDE LABOR

It's important to choose the right model for the job. Some image models like Midjourney are good at generating images in a specific style or with a certain aesthetic, while others offer advanced features like DALL-E's inpainting. Using multiple models together can expand the scope of what you can accomplish.

DALL-E has fewer features than most image models, but is great at this specific technique. It automatically blends the edges in so that the image fits well with the surroundings. As such you don't need to be particulary precise with the erase brush and you'll still get good results. What's remarkable is that you would struggle to get an image of the same quality using DALL-E 2 alone. Figure 8-17 shows what you would get if you prompted DALL-E with the same prompt used earlier for Midjourney.

Figure 8-17. Photograph of woman in a 1920s flapper party

DALL-E 3 provides superior quality, but at present is only available via API and in ChatGPT, not in the OpenAI labs interface used for inpainting. As image models proliferate, they are diverging in functionality and use cases, and it may take more than one model used in combination to accomplish the task you have at hand. Inpainting is a powerful technique for editing images, whether those images come from other AI models or a real life photographer.

Outpainting

Related to inpainting in DALL-E, is outpainting, where you generate outside of the frame of the existing image. This technique can in effect *zoom out* from the existing image to add context around it. This can be used to fill in more detail in an image you have generated or uploaded. It works

in the DALL-E editor just like inpainting, but you click *Add generation frame* to add another square, and type in your prompt for that square, as you can see in Figure 8-18, where legs have been added to the woman in the Van Gogh dress. Be sure to overlap with the existing image or the new frame will not resemble it. The greater the overlap the closer the frame matches the existing image, so for important images you may need to build up an image slowly and incrementally:

Figure 8-18. DALL-E Outpainting

Input:

```
van gogh style dress, inspired by starry night, blue and
yellow swirls, extremely detailed, very well lit, studio
light, 3.5 mm, dslr
```

The output is shown in Figure 8-19.

Figure 8-19. DALL-E Outpainting with extra frame

Often it can be difficult to achieve the right style purely with text prompts, particularly if the style is nuanced or you don't know all the words to describe it. Providing an example of an image to use for inpainting or outpainting is a shortcut to better results.

As well as creatively expanding on an existing image, outpainting is also helpful if you're trying to get an image in an aspect ratio other than square, by filling in the gaps. You keep adding new generative frames, prompting what you want in each new section of the image through trial and error until you find something consistent with the rest of the image, or until the full image is in the aspect ratio required (for example, going from portrait to landscape). This technique is also available as an extension in Stable Diffusion, but in our experience it's less reliable than DALL-E. In Midjourney this feature is called *zoom out*.

Consistent Characters

An underrated use of inpainting and outpainting is using an existing image to maintain consistency across generations. One such example is a common method for creating consistent characters by generating two images side by side, and inpainting one side at a time. First, generate an image while explicitly dictating that there are two images side-by-side, in a 2:1 aspect ratio:

Input:

```
two images side by side, rockstar American rough and ready
middle-age jawline actor man, photo booth portrait --ar 2:1
```

The output is shown in Figure 8-20.

Figure 8-20. Midjourney consistent character

The next step is to upscale one of the images, then mask one half of the upscaled image in an inpainting canvas, shown in Figure 8-21 using the Midjourney Vary Region feature.

Figure 8-21. Midjourney Vary Region

Finally, reprompt the masked part of the image using inpainting or Vary Region (as it's known in Midjourney), to dictate a different angle from the original portrait mode:

Input:

```
side profile image, rockstar American rough and ready
middle-age jawline actor man, photo booth portrait --ar 2:1
```

The output is shown in Figure 8-22.

Figure 8-22. Consistent characters in side profile

This inpainting and generation process can be repeated for multiple an-
gles, with the express purpose of finding new images of a character that
looks identical to the original one you generated. Because one half of the
image is always present, the model maintains consistency of the charac-
ter's features across generations, allowing you to build a more compre-
hensive perspective of a single character in different poses and positions.
All that you need to do to create an image of the character in a new situa-
tion is inpaint half of the 2:1 image with the new prompt, and crop it in
Photoshop (or some equivalent).

PROVIDE EXAMPLES

Many people think of using a real image as a baseline when prompting for in-
painting, but many of the more advanced AI artists use generated images them-
selves as inputs to maintain control over the consistency of the characters or ob-
jects in their story.

Prompt Re-writing

One of the issues you may run into when putting an AI system into production is that you can't expect the users of your system to be expert prompt engineers. It's a case of garbage in, garbage out: if they write a substandard prompt they'll get poor results and complain about the quality of your product. One common trick in the industry is to rewrite the prompt to make it better and more likely to get impressive results. This is a form of *meta prompting* where the prompt for one AI model is written by another.

Imagine a simple application where a user inputs a subject and an artist, then an image is generated of the subject in the style of the artist. The prompt template is `a {subject} in the style of {artist}`:

Input:

```
a daschund dog in the style of banksy
```

The output is shown in Figure 8-23.

The issue with this prompt is that the expectation would be that the dog would be part of the street painting (in Banksy style), whereas instead it is standing next to it in the image that was generated. To fix this, you can take the user prompt, and inject that into a prompt to ChatGPT to find the artist's medium:

Input:

```
What's the medium that the artist banksy mostly used? Respond
in 1-3 words only.
```

Output:

```
Street art
```

Finally, you can use this output to rewrite the original user prompt, in the format `{medium} of a {subject} in the style of {artist}`:

Input:

```
street art of a daschund dog in the style of banksy
```

The output is shown in Figure 8-24.

Figure 8-24. Street art of a daschund dog in the style of banksy

This system can be built out further to include other prompt engineering techniques, such as quality boosters or negative prompts, to make the results more reliable. It's possible to get good results just asking ChatGPT to rewrite the prompt for DALL-E (which ChatGPT Plus has available as a tool) and then use what it gives you for other models. Some attempts have been made to train AI models that specialize in generating high-quality

prompts, though in our experience this method only brings quality up to average, and doesn't beat expert prompt engineering.

Instead of expecting a non-technical person to submit a good quality prompt, simply pass their input to another AI model that can help improve the original prompt.

Meme Unbundling

The main issue with replicating an artist's style, or emulating an art movement, is that it's relatively unoriginal. Nobody knows what the legal implications of AI art will be, but certainly artists like Greg Rutkowski and others have already spoken out about the immorality of copying their style.

One way to still get the benefit of the shortcut of referring to a successful artist or art movement, while being more original in your style, is to unbundle the memes of an artist. The word meme doesn't just mean *a funny viral image*, it refers to any piece of cultural information that gets copied from person to person. If you can decompose an art style into its component parts and characteristics, then you can use these sub-components to remix your own style. Say, for example, you were trying to make a painting of Times Square in the style of Salvador Dali:

Input:

```
a painting of times square in the style of "The Persistance
of Memory" by Salvador Dali
```

The output is shown in Figure 8-25.

Figure 8-25. Painting of Times Square in the style of "The Persistance of Memory" by Salvador Dali

You are already getting good results by evoking Dali's name, but it's not quite right. You want a surrealist style, but not an exact copy of Dali's work. The term *Unbundling* was coined in this context by Bakz T. Future, meaning to extract the characteristics of an artist's style, in order to get something familiar but unique. It works by asking a text generation model (in this case ChatGPT-4) to describe the style of the artist you're emulating, which you can then use as a prompt:

Input:

```
Briefly describe the characteristics and artistic style of
"The Persistance of Memory" by Salvador Dali, without
mentioning the artwork or the artist:
```

Output:

```
This is a notable example of Surrealist art, steeped in
dream-like, fantastical imagery that pushes the boundaries of
traditional understanding and interpretation. The painting is
composed with a distinctly precise and detailed technique,
reflecting a deceptive realism.

...

Contrasting with the melting objects, there are other items
in the scene that exhibit solidity and permanence, creating
an uncanny juxtaposition. The painting's composition and
imagery invite deep exploration and various interpretations,
concerning themes of reality, dreams, time, and the
subconscious. Overall, it is a visually striking work that
embodies the ethos of its movement: challenging the viewer's
perceptions and exploring the unconventional and the
irrational.
```

You must make sure the prompt insists on not mentioning the artists or the artwork, otherwise it won't work as well as a prompt. You want to describe the style to someone who has never seen it, so that the description is clean. For reference Figure 8-26 shows the famous painting ChatGPT is describing.

Now that you have a description of one of Dali's most iconic paintings, you can understand what makes it so visually appealing. In particular, this sort of explanation can be useful if you never went to art school, and otherwise wouldn't know how to describe these elements. From here you need to abbreviate the description to fit into an image model prompt:

Input:

```
Convert this description into a shortened prompt for an AI
image generation model such as DALL-E:
```

Output:

```
Surrealist landscape with subdued colors of brown, blue, and
yellow. Depict a barren, craggy shoreline and calm sea
meeting a clear sky. Include amorphous, melting objects
resembling common items, indicating fluidity and impermanence
of time. Contrast these with other solid, permanent elements.
Evoke a sense of stillness, dreamlike atmosphere, and
quietude.
```

Take this output and add the subject of your painting, Times Square. It can also help to modify the prompt to make it flow better, as the prompts ChatGPT writes can be too instructive:

Input:

```
Painting of times square, surrealist landscape with subdued
colors of brown, blue, and yellow, a barren, craggy shoreline
and calm sea meeting a clear sky. Include amorphous, melting
objects resembling common items, indicating fluidity and
impermanence of time. Contrast these with other solid,
permanent elements. Evoke a sense of stillness, dreamlike
atmosphere, and quietude.
```

The output is shown in Figure 8-27.

Figure 8-27. Unbundled Dali memes applied to a painting of Times Square

Rather than guiding the AI image model towards a specific artist's work, you can emulate a close approximation by using a description of the artist's work. This is more transformative and creative approach than simply evoking an artist's name, and perhaps more ethical.

This image is still similar to Dali's work, but it has been transformed through the filter of ChatGPT's description. Therefore it's already more original than what you got when you simply evoked his name: an advantage over the average prompter. But you're in an even better position now, because you have unbundled Dali's style into individual *memes* like "surrealist landscape", "melting objects", and "dreamlike atmosphere", and can more easily remix its component parts to make the image more unique:

Input:

```
Painting of times square, surrealist landscape with subdued
colors of orange, red, and green, imposing buildings and calm
river meeting a stormy sky. The amorphous melting dripping
clock in the center of the square indicates the fluidity and
impermanence of time in contrast with other solid, permanent
elements. Evoke a sense of stillness, dreamlike atmosphere,
and quietude.
```

The output is shown in Figure 8-28.

You have only made small modifications to the color and elements in the painting, but you could go further. It's also possible to take elements from other popular artists and combine the aspects you like to arrive at something new. This technique only works right now with artists and artworks that are famous enough to be readily described from the training data, however as AI models become multi-modal (i.e. able to generate both images and text), expect to be able to feed in an image and get a description to use for unbundling.

Meme Mapping

One of the most common forms of prompt inspiration is looking at what prompts other prompt engineers are getting results with. The Midjourney Discord community has millions of active members, with thousands of

new images being generated and automatically shared every day, as do other AI communities including on Reddit and across various other websites, email newsletters, and social media accounts. One commonly used website is Lexica.art, which has a searchable database (by keyword and similarity) of many Stable iffusion images and their prompts for inspiration.

While searching and browsing through these sources of inspiration, you're likely to notice recurring patterns, or memes, in the words that are used for a particular type of image. call this process of intentionally and systematically finding these patterns meme mapping, and it can be an invaluable tool for identifying useful prompts. For example, you may search `super mario` on Lexica and see lots of examples where people have tried to create a *realistic* Mario, like the one in Figure 8-29, which might inspire you to do the same, starting with a prompt that's already proven to work, saving you considerable time.

Figure 8-29. Realistic Mario

Alternatively you might apply this meme to a character from a different franchise, and try repurpose some of the prompts used by others to get a realistic effect. Without doing this research you might not have been aware image models could generate real-world versions of cartoon or game characters, or perhaps never would have thought to try it. You may have never stumbled upon the insight that including `as a soviet factory worker` in your prompt helps evoke a sense of gritty realism,

and may never have encountered the work of the two artists referenced. There is a healthy culture of remixing content in the AI art community, with people learning from other's prompts, and then paying it forward by sharing their own expertise:

Input:

```
portrait of homer simpson as a soviet factory worker, gritty,
dirty, beautiful, very detailed, hyperrealistic, medium shot,
very detailed painting by Glenn Fabry, by Joao Ruas --no
cartoon
```

The output is shown in Figure 8-30.

Figure 8-30. Realistic Homer Simpson

This meme mapping process can be done manually, with the examples copied and pasted into a spreadsheet or productivity tool Notion, although that can be time consuming. So long as you are respecting a web-

site's terms and conditions and any legal obligations in your country, it would also be possible to write custom code to programmatically scrape the contents of that website. Once you have all the data in one place, you could programmatically label the images with an entity recognition model like Google Vision, a multimodal model like GPT-4 Vision, or use Natural Language Processing (NLP) such as NGRAMs analysis on the prompts in order to identify patterns at larger scale than is possible manually.

Prompt Analysis

One common mistake is to continue to build out longer and longer prompts, without thinking about what parts of the prompt are really necessary. Every word added perturbs the model in some way, adding noise to the resulting output. Often unnecessarily removing words can be as effective as adding new words. To conduct this analysis without lots of trial and error, Midjourney offers a `/shorten` command which attempts to remove these unnecessary words, leaving only the core tokens which the model pays the most attention to. Click show details at the bottom of the response to get token-level weightings and a visual chart:

Input:

```
portrait of homer simpson as a soviet factory worker, gritty,
dirty, beautiful, very detailed, hyperrealistic, medium shot,
very detailed painting by Glenn Fabry, by Joao Ruas
--no cartoon
```

Output:

```
**portrait** (0.08) of **homer simpson** (1.00) as a
**soviet** (0.19) **factory** (0.21) **worker** (0.08),
gritty (0.02), dirty (0.02), beautiful (0.00), very (0.00)
detailed (0.01), hyperrealistic (0.01), medium (0.00) shot
(0.00), very (0.00) detailed (0.01) painting (0.05) by Glenn
Fabry (0.08), by **Joao Ruas** (0.09)
```

```
████████████  homer simpson
█░░░░░░░░░░░  factory
█░░░░░░░░░░░  soviet
█░░░░░░░░░░░  portrait
█░░░░░░░░░░░  worker
█░░░░░░░░░░░  Joao Ruas
█░░░░░░░░░░░  Glenn Fabry
█░░░░░░░░░░░  painting
```

Once you have this analysis, you can use it to cut any noise from the prompt and zero-in on what words, or memes, are actually important to the final result.

EVALUATE QUALITY

Seeing the weights the model assigns to each token gives you unparralleled insight into how the model works. Often we make assumptions about what's important in a prompt, and those assumptions can be quite far from reality.

Summary

In this chapter, you have learned about standard practices for image generation using diffusion models. You explored format modifiers such as stock photos, oil paintings, and Egyptian hieroglyphs, and how they can be used to create unique and visually appealing images. Additionally, you discovered art style modifiers that allow for the replication of popular art styles or artists, such as Lewis Carroll's *Alice in Wonderland style*.

You delved deeper into the application of prompt engineering principles. You learned how to use art style modifiers to replicate popular art styles and artists, and how mentioning specific artists' names can help achieve the desired visual style. The concept of negative prompts and weighted terms was introduced, allowing you to specify what you don't want in an image and control the mixture of different concepts. You also explored the concepts of inpainting and outpainting, where specific parts of an image can be generated separately by erasing and adding prompts. You dis-

covered how these techniques can be further expanded and combined to enhance the reliability and quality of generative AI results.

In the next chapter, you will dive deeper into the world of image generation and explore more advanced use cases. You will learn how to harness the power of Stable Diffusion and AUTOMATIC1111 to improve your image generation skills. Including advanced Stable Diffusion techniques like utilizing ControlNet models for more control over the style and composition of your images, you will discover a wide range of exciting possibilities.