Image generation with Stable Diffusion

- Apoorv Sharma

1. Model uploaded on github -

https://github.com/CrunchyExplosion/StableDiffusion_imageGenerator

2. Can we generate images of different aspect ratios using Stable Diffusion?

Yes, Stable Diffusion (SD) can generate images of different aspect ratios. By specifying different width and height dimensions during the generation process, you can adjust the aspect ratio of the output images.

The height and width arguments in the image generation function allow you to set specific dimensions, thus altering the aspect ratio.
For example:

- Aspect ratio 1:1: Set both height=512 and width=512 for a square image.
- Aspect ratio 2:1: Set height=512 and width=1024 for a wider rectangular image.

Comment on Image Generation Quality:

- **Aspect Ratio 1:1**: When generating images with a 1:1 aspect ratio (512x512), the image quality tends to be consistent and sharp. Since the Stable Diffusion model is often optimized for this square aspect ratio, images generally appear well-structured, with balanced compositions.
- **Non-Square Aspect Ratios**: For aspect ratios other than 1:1 (e.g., 2:1 or 3:2), the generation quality may vary depending on the dimensions and the prompt. The model stretches or compresses the image to fit the dimensions, which can sometimes lead to slight distortions, loss of detail, or unnatural compositions in areas of the image. The depth map helps preserve structural integrity to some extent, but non-square ratios may still show minor artifacts.

In general, Stable Diffusion adapts well to different aspect ratios but may need further adjustments for the best results when working with extreme aspect ratios.

3. What is the generation latency? How can it be reduced?

Generation Latency Measurement:
The generation latency refers to the time it takes for the model to create an image. Latency can be measured using Python's time.time() function before and after the image generation step. For example, generating an image with 50 inference steps and a 512x512 resolution typically takes around **20-30 seconds** on a CPU, depending on the hardware configuration.

Quick Fixes to Reduce Latency:

- **Reduce the number of inference steps**: By default, Stable Diffusion uses 50 inference steps. Reducing this number to 25 (or even lower) can significantly cut down the generation time while still maintaining acceptable image quality. For instance, reducing the steps from 50 to 25 can cut the latency by **up to 50%**.

Example reduction:

  - 50 steps: ~30 seconds
  - 25 steps: ~15 seconds

- **Use a GPU for faster computation**: If you are currently running the model on a CPU, switching to a GPU can drastically reduce the generation time. On a GPU like an NVIDIA A100 or V100, latency can drop to **2-5 seconds** for a similar task.

- **Optimize the model**: There are various optimizations, like using half-precision floating point (fp16) to reduce memory usage and computation time, which can further decrease latency.

Comment on Reduced Latency and Generation Quality:

- **Reduced Inference Steps**: When lowering the inference steps from 50 to 25, you may observe a slight decrease in image quality, particularly in terms of sharpness and detail. However, the drop in quality is often minimal and acceptable for many use cases, especially when prioritizing speed over high precision.
- **GPU Acceleration**: Using a GPU massively reduces the time to generate an image while maintaining the same level of quality as on a CPU. This is the most effective way to cut latency without sacrificing quality.