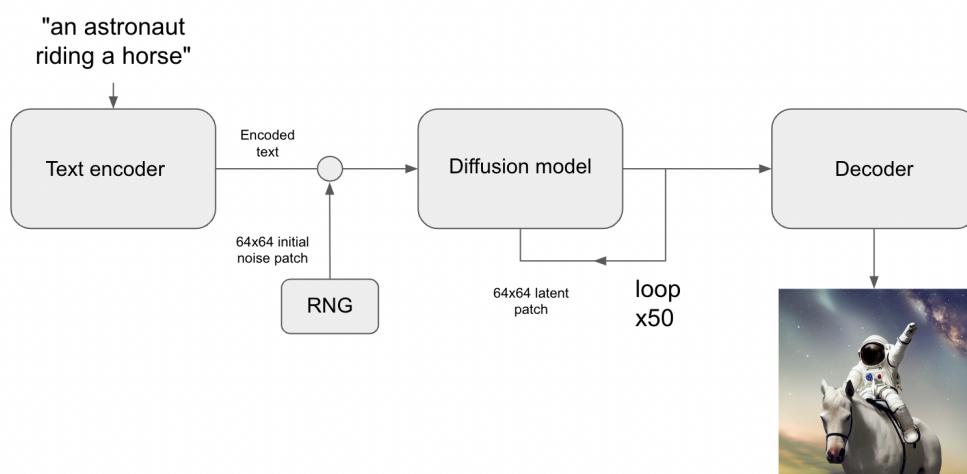
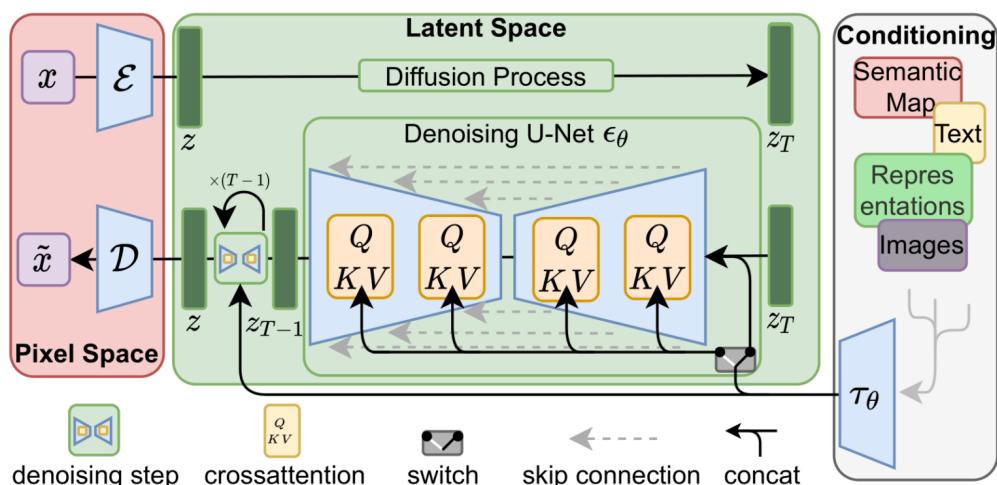


Este documento fornece uma introdução sucinta às inteligências artificiais geradoras de imagens, focando especificamente no Stable Diffusion. Ele explica brevemente o funcionamento dessa tecnologia e apresenta opções para sua utilização. As opções são divididas em: executar localmente, executar em jupyter notebooks em plataformas como o Google Colab ou utilizar plataformas digitais. Por fim, são fornecidas algumas dicas sobre como otimizar os resultados do Stable Diffusion.

## Stable Diffusion

Implementa um modelo de difusão latente, uma espécie de rede neural gerativa profunda. Tanto a implementação como os pesos do modelo foram lançados como código aberto. Uma explicação do seu funcionamento pode ser encontrada [aqui](#).



## Rodar localmente

[Automatic1111](#) - A GUI mais utilizada para executar o Stable Diffusion, que possui opções como txt2img, img2img, inpainting, ControlNet e muito mais, é recomendada para sistemas com pelo menos 4GB de VRAM e 8GB de RAM. Esse [artigo](#) apresenta as principais funcionalidades.

No caso de problemas de memória na GPU, você pode tentar as seguintes soluções:

- Substitua o arquivo "webui-user.bat" pelo disponibilizado no repositório (Versão medvram e lowvram(mais lento com menor uso de vram)).
- Execute o "auto1111" por meio desse arquivo.
- Nas configurações, vá para a seção de otimização e selecione a opção "xformers".
- Reinicie o "auto1111".

[Easy Diffusion 2.5](#) - Uma GUI que possui a maioria das opções do "auto1111" e funciona bem com 4GB de VRAM (usando a configuração low nas opções) e pelo menos 8GB de RAM, capaz de gerar imagens de até 768x768. Essa GUI permite a execução em CPU, mas pode ser mais lenta (não foi testado nessa configuração). Além disso, o recurso de inpainting funciona bem nessa GUI. É possível carregar modelos e plugins nela para ampliar suas funcionalidades. Possui opções prontas para criar prompts mais detalhados.

## Rodar remotamente

Antes de começar, é necessário escolher uma plataforma para executar os Jupyter Notebooks. A opção mais simples é utilizar o Google Colab, com um ambiente de execução que possua suporte a GPU. Além do Google Colab, existem outras plataformas disponíveis, como o Paper Space e o Vast AI, que também podem ser utilizadas. No entanto, as opções que serão apresentadas a seguir e que estão disponíveis no repositório foram testadas especificamente no Google Colab.

Automatic1111 Collab - Roda automatic1111 pelo google collab. Rápido de configurar e rodar. Podem ocorrer problemas de desconexão no

Google Colab sem versão paga e é necessário baixar modelos no Google Drive.

Stable Diffusion Exemplo - Notebook exemplificando o uso de stable diffusion para implementações próprias para quem tem interesse na API do stable diffusion. Pode ser rodado com 6GB de VRAM.

Stable Diffusion Notebook - Feito para rodar pelo google collab, sendo rápido e fácil de configurar. É possível carregar outros modelos nele a partir do site [hugging face](#) colando diretório do repositório. Não é necessário armazenar modelos no Google Drive. Permite acesso de txt2img e configurações básicas do stable diffusion.

Para as próximas duas opções é necessário pesquisar como configurar para rodar (Provavelmente há templates prontas que podem ser usadas).

Paperspace - oferece uma opção de GPU gratuita, mas é muito disputada. Não foi possível testar qualidade. Há também uma opção de \$8 (R\$40) por mês para acesso ilimitado a boas GPUs.

Vast AI - oferece serviços para alugar GPU. Preço razoável (\$0,402/hr).

## Plataformas Online

[Dreamlike](#) - possui limitação de créditos, com um máximo de 24 por mês e 50 no total na versão gratuita. Possui alguns modelos disponíveis para realismo e arte digital. Oferece opções prontas para prompt negativo. "img2img" não é muito bom.

[Scenario](#) - Plataforma focada em geração de assets para games com ia. Muitas opções de modelos para diferentes usos (Personagens, objetos, paisagens, etc)

## Como melhorar resultados do Stable Diffusion

Para se obter bons resultados com o Stable Diffusion é necessário entender como escrever bons prompts para ele. Uma explicação mais completa de como escrever bons prompts pode ser encontrada [aqui](#), mas a seguir está um resumo.

Em geral, é bom lembrar do seguinte: Stable Diffusion não lê mentes. Quanto mais específico o prompt, melhor será o resultado. Se informarmos como prompt “dog” o resultado será um cachorro, mas a variação da imagem será grande.



Agora, se informarmos um prompt como “dog running on a grass field” o resultado será muito mais específico.



## 1. Anatomia do prompt

Para prompts pode-se dividir as palavras chaves nas seguintes categorias:

- Subject - Sujeito ou objeto principal da imagem. Quanto mais detalhes melhor será o resultado.
- Medium - Material usado para produzir imagem (Photography, Digital art, Painting)
- Style - Estilo artístico da imagem
- Artist - Usa estilo de um artista como referência
- Website - Usa resultados de um website como referência
- Resolution - Palavras chaves que ajudam a melhorar a qualidade e nitidez da imagem
- Additional details - Adjetivos e características que ajudam a especificar a imagem
- Color - Ajuda a controlar a cor das imagens que pode se manifestar nos tons da imagem ou nos objetos em si
- Lighting - Iluminação é muito importante na composição de imagens e portanto pode ter grande influência no resultado final.

É importante ressaltar que nem sempre o efeito será muito aparente para adição de uma palavra ou outra e que não há a necessidade de sempre se colocar todas as categorias no prompt. Uma regra geral é adicionar um prompt simples com sujeito e meio e ir gerando imagens e adicionando mais informações no prompt conforme necessário.

Exemplos de palavras podem ser encontradas [aqui](#)

## 2. Prompt negativo

No stable diffusion é possível adicionar um prompt negativo que informa o que é indesejado na imagem. Isso pode ser tanto coisas concretas como coisas mais abstratas. Um prompt universal que pode ser utilizado é o seguinte:

ugly, tiling, poorly drawn hands, poorly drawn feet, poorly drawn face, out of frame, extra limbs, disfigured, deformed, body out of frame, bad anatomy, watermark, signature, cut off, low contrast, underexposed, overexposed, bad art, beginner, amateur, distorted face, blurry, draft, grainy, title, text

### **3. Peso nas palavras chaves**

O peso de uma palavra no prompt pode ser ajustado com a seguinte sintaxe: (keyword:factor). Por exemplo, no prompt “(dog:1.2) running on grass field” aumentamos a influência do conceito de um cachorro na imagem. Uma sintaxe alternativa é simplesmente utilizar parênteses que aumentam por uma fator de 1.1 ou colchetes que diminuem por um fator de 0.9. Esses fatores são cumulativos da seguinte forma:

$$\begin{aligned}(\text{dog}) &= (\text{dog}:1.1) \\ ((\text{dog})) &= (\text{dog}:1.21) \\ (((\text{dog}))) &= (\text{dog}:1.33)\end{aligned}$$

$$\begin{aligned}[\text{dog}] &= (\text{dog}:0.9) \\ [[\text{dog}]] &= (\text{dog}:0.81) \\ [[[dog]]] &= (\text{dog}:0.73)\end{aligned}$$

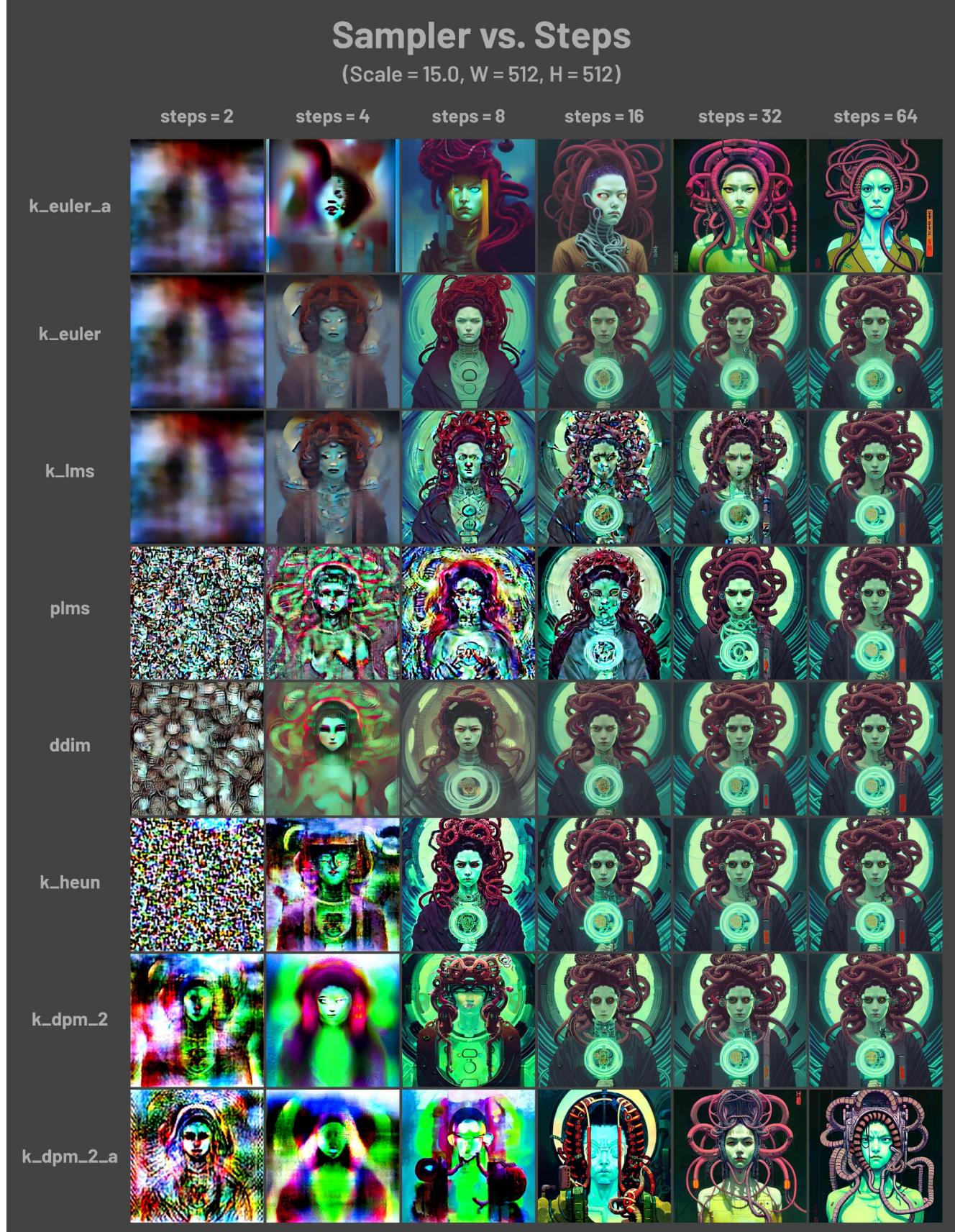
Outro método é a combinação de palavras chaves, explicado no link referenciado anteriormente.

### **4. Modelo utilizado**

Por se tratar de uma modelo open-source, a comunidade do Stable Diffusion criou centenas de modelos personalizados. Estes possuem forte influência nos resultados obtidos por um prompt e podem ajudar muito quando se procura gerar um tipo específico de imagem (disney, anime, hq, pintura, fotografia, etc.). Dois sites que podem ser visitados para encontrar modelos são o [hugging face](#) e o [civtai](#).

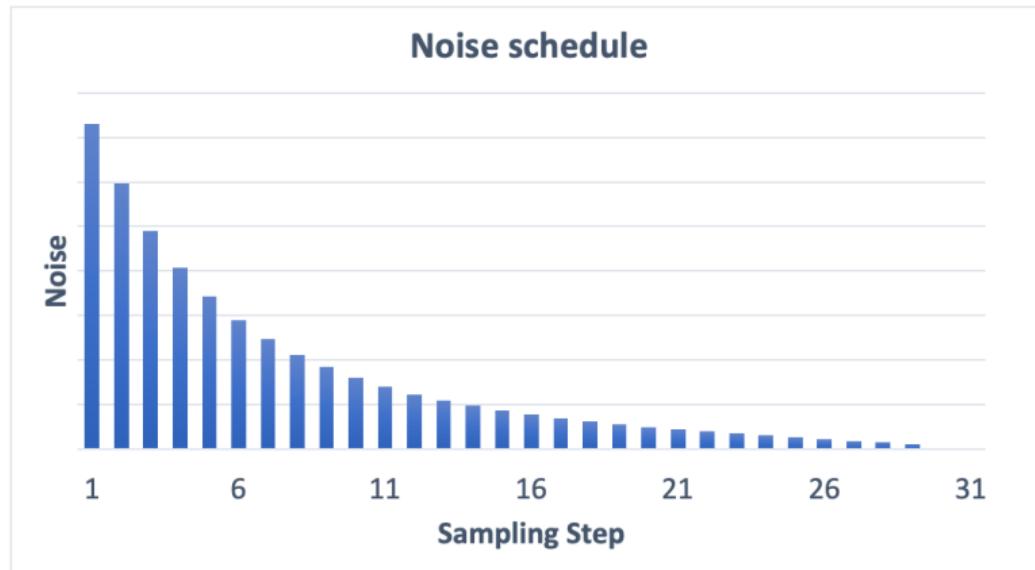
## 5. Sampler

Os samples determinam o ruído em cada steps da geração e portanto influenciam fortemente no resultado final. Uma explicação completa e comparação entre os diferentes samplers pode ser encontrada [aqui](#). Os diferentes resultados podem ser observados a seguir:



## 6. Quantidade de steps

A quantidade de ruído na imagem diminui conforme se realiza os steps. Por isso, para samplers que convergem os primeiros steps são os mais importantes para a definição dos resultados. Após isso, os steps configuram mais um processo de aperfeiçoamento da imagem. O valor recomendado varia dependendo do sampler utilizado. Em geral, um valor de 20 a 30 sampling steps é suficiente para uma ótima imagem.



## Outro técnicas interessantes

Inpainting - Uma explicação de como utilizar inpainting pode ser encontrada [aqui](#).

Controlnet - A melhor opção para executar o ControlNet é pelo automatic1111. A explicação completa de como fazer isso pode ser encontrada [aqui](#). Os modelos otimizados do ControlNet (necessários para rodar com 4GB de VRAM) podem ser encontrados [aqui](#). É importante ressaltar que dependendo dos modelos utilizados, podem ocorrer problemas de memória, por isso é recomendado verificar essa questão. Uma opção é utilizar o “webui-user.bat” para lowvram disponível no repositório.

## Lora

Uma forma de adicionar informação sobre um modelo existente. Uma explicação sobre lora se encontra [aqui](#)