EDM4611 – Ressource “Text to Image”

« Text to image » ou « image to image »

Les systèmes de texte à images sont des outils qui utilisent l’intelligence artificielle pour créer des images à partir de descriptions textuelles. Ils fonctionnent en analysant le texte et en générant des représentations visuelles correspondantes. L’IA dans les systèmes de texte à image utilise généralement des réseaux de neurones profonds pour apprendre à associer des mots à des pixels. Il existe plusieurs modèles d’IA qui peuvent réaliser cette tâche, comme les réseaux génératifs adverses (GAN), les réseaux de diffusion, ou les réseaux de transformation. Ces modèles sont entraînés sur de grandes bases de données d’images et de textes, et apprennent à extraire les caractéristiques visuelles pertinentes à partir des descriptions textuelles. Ensuite, ils utilisent ces caractéristiques pour générer des images synthétiques qui correspondent au texte.

Les réseaux de neurones

Un réseau de neurones profond est un type d'intelligence artificielle qui utilise des couches de calculs pour apprendre à partir de données complexes. Un réseau de neurones profond est composé d'au moins deux couches cachées de neurones artificiels, qui sont des unités de traitement représentés par des formules mathématiques simples. Les neurones artificiels sont connectés entre eux par une fonction de « weights » ou le poids de connexion. L’apprentissage dans ses systèmes se fait par l’ajustement automatique des poids qui lient chaque neurone afin de faire correspondre une donnée d’entrée à une donnée de sortie. Chaque couche reçoit des informations de la couche précédente, effectue une transformation non linéaire, et transmet le résultat à la couche suivante. La première couche est appelée couche d'entrée, et la dernière couche est appelée couche de sortie. La couche de sortie produit la réponse finale du réseau, qui peut être une prédiction, une classification, une génération, ou toute autre tâche. Les couches intermédiaires sont appelées couches cachées, car elles ne sont pas directement observables. Elles permettent au réseau d'apprendre des caractéristiques abstraites et hiérarchiques des données, qui sont utiles pour résoudre des problèmes complexes.

En savoir plus :

1. https://datascientest.com/fonctionnement-des-reseaux-neurones
2. https://datascientest.com/deep-neural-network

Stable Diffusion

Pourquoi Stable Diffusion

1. Modèle open source
2. Accessible par différent moyen
    a. En ligne (ressources gratuites)
    b. En ligne (ressources payantes pour plus de rapidité)
    c. En local (ressources gratuites, sur un grand nombre de GPU)
3. Un grand nombre de modèles disponibles
4. La puissance de SDXL 1.0 et l’inclusion de « Refiners » et « Lora »

5. La flexibilité de Automatic1111 pour le contrôle des « prompts »

Accéder à Stable Diffusion

1. En ligne
    a. https://clipdrop.co/stable-diffusion
    b. https://dreamstudio.ai/
2. En local
    a. Automatic1111 https://github.com/AUTOMATIC1111/stable-diffusion-webui
    b. Comfy UI https://github.com/comfyanonymous/ComfyUI

Accéder à de nouveaux modèles et Lora

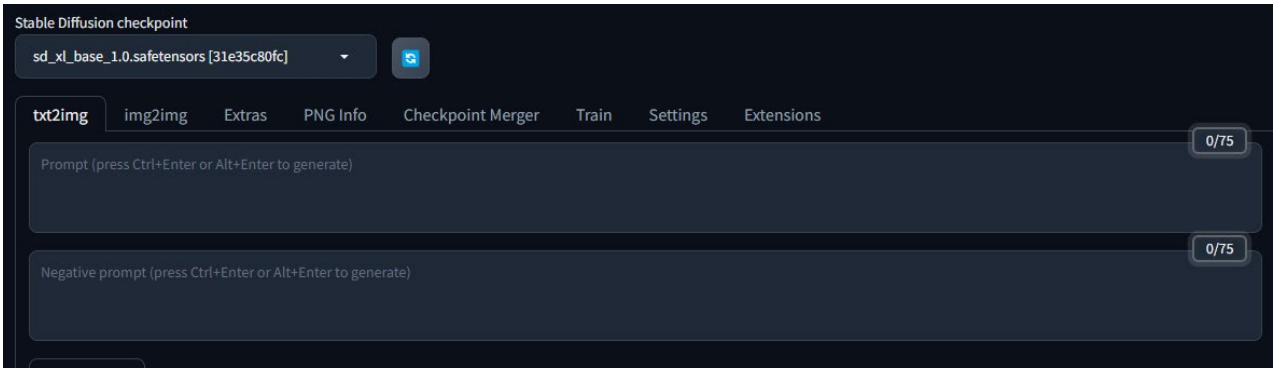
Télécharger des modèles de la communauté sur https://civitai.com/

L'interaction en différent modèles

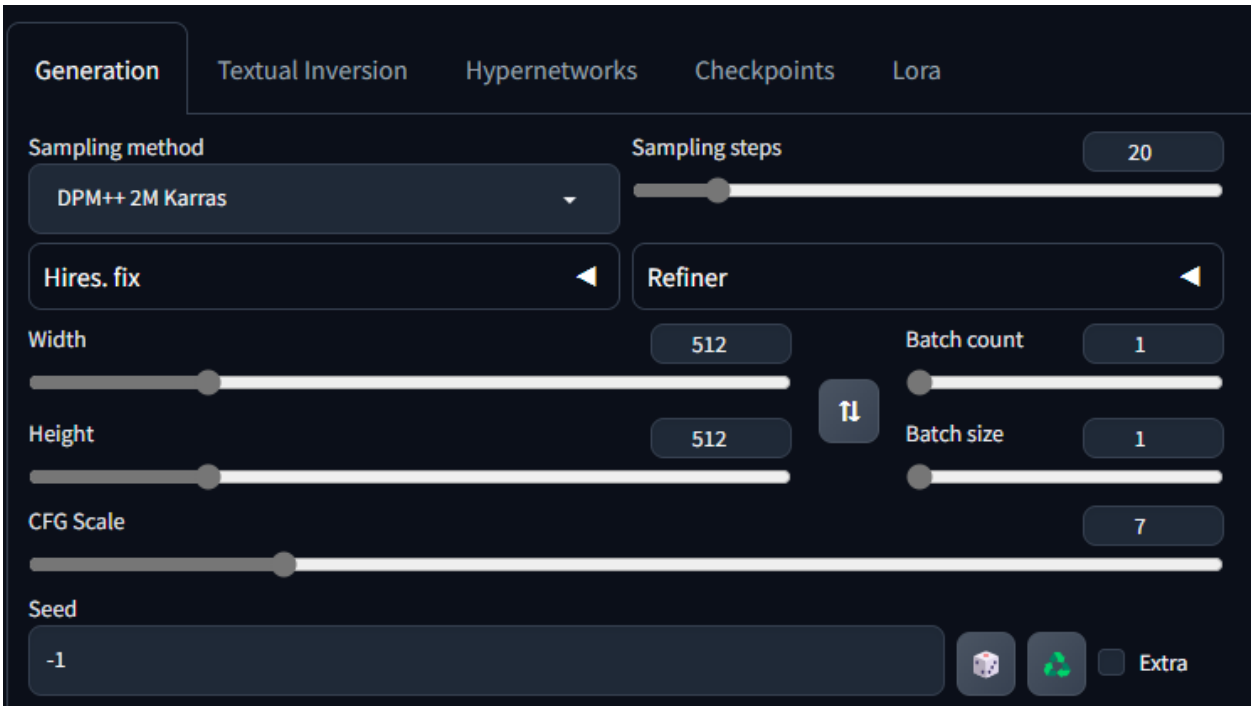
1. Modèle de base
    a. Nous allons utiliser SDXL 1.0 comme modèle de base pour comprendre le prompt et générer la forme de base de l'image représentant la sémantique du texte
2. Refiner
    a. Nous allons utiliser le SDXL 1.0 Refiner pour raffiner les détails de l'image. Ce modèle accentue la précision des visages et des mains ainsi que la cohérence des objets partiellement masqués
3. Lora
    a. Un Lora est un type de modèle qui permet d'ajuster la sortie d'un modèle de base de Stable Diffusion selon un concept ou un thème spécifique, comme un style artistique, un personnage, une personne réelle, ou un objet. Un Lora est généralement beaucoup plus petit qu'un modèle de base, et s'applique comme une couche supplémentaire de transformation.

Les paramètres dans Automatic1111

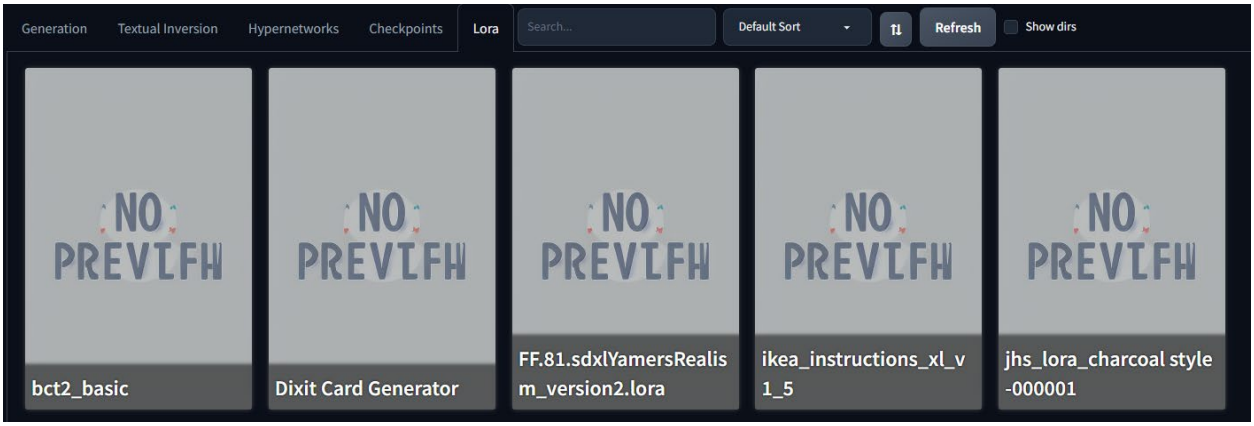
1. Le modèle principal (checkpoint), la prompt et la prompt négative
    a. Le modèle peut être sélectionné à partir du menu « Checkpoint » on peut tenter de générer la même prompt avec plusieurs modèles pour comparer les résultats
    b. La prompt de base est ce qu'on cherche à représenter dans l'image
    c. La prompt négative est ce qu'on veut exclure de l'image.



2. Les paramètres de base
    a. Sampling method
        i. L'algorithme utiliser pour échantillonner
    b. Sampling steps
        i. Le nombre de passage que le système fera pour générer votre image. Généralement +grand = meilleure qualité = plus long temps de génération
    c. Hires. Fix
        i. Augmenter la résolution de l'image (upscaler)
    d. Refiner
        i. Sélection du modèle de raffinement et de l'étape ou le modèle commence à affecter la génération de l'image
    e. Width / Height
        i. La taille de l'image (attention augmenter la taille augmente grandement le temps de génération)
    f. Batch count / Batch size
        i. Combien d'images seront générées
    g. CFG Scale
        i. Comment le modèle se conforme au prompt. Plus petit = plus créatif
    h. Seed
        i. Le « random seed » utilisé pour la génération. Il y a possibilité de contrôler le seed afin de créer des grandes ou petites variations dans les images générées à partir de la même prompt



3. Sélection du Lora (optionnel) pour ajuster le style de l'image



Construire vos prompts

Stable Diffusion permet une incroyable flexibilité et précision dans la création de prompt. Il est possible de créer des prompts complexes avec plusieurs détails voulu pour piloter le modèle dans la direction souhaiter.

Attention : pour de meilleurs résultats focusser la prompts sur un sujet et donner des détails à propos de ce sujet.

Un format de prompt pour obtenir des résultats photographiques (aucun Lora)

[STYLE OF PHOTO][1] photo of a [SUBJECT][2], [IMPORTANT FEATURE][3], [MORE DETAILS], [POSE OR ACTION][4], [FRAMING][5], [SETTING/BACKGROUND][6], [LIGHTING][7], [CAMERA ANGLE][8], [CAMERA PROPERTIES][9], in style of [PHOTOGRAPHER][10]

Source : https://promptgeek.gumroad.com/l/photoreal

*Utiliser l'anglais

** Pour chaque mot clé il est possible d'ajuster le poids en utilisant le format : (mot-clé:1.5)

1. Le style de la photo souhaiter (par exemple : « documentary », « lifestyle », « abstract », « analogue », « instant photo »)
2. Le sujet principal de la photo
3. Quelques caractéristiques importantes, chacun séparé par des virgules
4. L'action du sujet, utiliser un verbe d'action « kicking » « dancing » « spinning »
5. Le cadrage de la photo « close shot » « upper body » « full body »
6. Dans quel environnement se trouve votre sujet?
7. Quel est l'éclairage de votre photo? « soft lighting » « golden hour » « overcast » « creative shadow play »
8. L'angle de la caméra (comment la photo est elle prise?)
9. Le type de caméra ou les propriétés de la caméra
10. Optionnel : Dans le style d'un artiste

Exemple :

<lora:sd_xl_offset_example-lora_1.0:1> An analog photo of a business man, wearing dark clothes and sunglasses, drinking tea, upper body shot, a cavern in the background with dripping water, creative shadow play, eye level angle, shot on red camera

<lora:sd_xl_offset_example-lora_1.0:1> An (documentary:1.6) photo of a baobab tree surrounded by birds, large trunk, lush green leaves, large shot, in a clear meadow, golden hour lighting, shot on Kodak Vision3

Styles et cohérence

Il est possible de construire vos prompts (positives et négatives) en ayant en tête un certain style et de changer le sujet en gardant vos mots-clés de style identiques afin de créer des images cohérentes dans un style particulier

Il existe un ensemble de style créés par la communauté desquels vous pouvez vous inspirer pour créer vos propres styles.

Voir la feuille de style XLSX dans le Github

Si vous utiliser Automatic1111 en local, vous pouvez utiliser un gestionnaire de styles (https://github.com/ahgsql/StyleSelectorXL) en extension afin de facilement gérer les styles en créant des boutons cliquables pour ajouter un style.