

L'intelligence artificielle pour les rendus d'architecture

L'impact des différents paramètres sur le style graphique

Master séminaire : SAPI
ENSAPLV

Réalisé par

MAHDI El Khalil Abdeldjalil

Encadrant :

- Joaquim Silvestre
- Anne Tuscher

2024/2026

Soutenance le 13/01/2026

Dédicaces :

Ce mémoire est l'aboutissement d'un travail intense, rendu possible grâce au soutien indéfectible de ceux qui me sont chers.

- Je dédie ce travail avant tout à mes parents, ma sœur et mon frère. Bien que la distance nous sépare, votre présence dans mon cœur est constante. Merci de croire en moi, de m'encourager sans relâche et de me donner la force de franchir chaque étape.
- À mes amis, Nissou, Raynou et Mentaaliti, et tous les autres pour leur "harcèlement" affectueux chaque vacance pour qu'on se retrouve au rond-point ; ces moments de déconnexion ont été mon équilibre.
- Et à ma punaise, qui m'accompagne et me soutient chaque jour malgré la distance.
- Je dédie aussi ce travail à mes collègues de travail, mes coéquipiers de basket et les enfants que je coach, et surtout à mon coach de basket cherif.

Remerciements

- Je tiens tout d'abord à exprimer ma profonde gratitude envers Dieu pour m'avoir guidé et donné la force nécessaire à l'accomplissement de ce travail.
- Mes remerciements s'adressent ensuite à mon équipe encadrante, Madame Anne Tüscher-Dokic et Monsieur Joaquim Silvestre, pour la qualité de leur suivi et le cadre pédagogique stimulant qu'ils ont su instaurer.
- Ma reconnaissance va aussi aux étudiants ayant participé à mon expérimentation. Merci pour votre implication, votre curiosité et votre aide durant les séances d'apprentissage.
- Enfin, je remercie chaleureusement mon ami abdennafa, qui travaille dans le data science et qui m'a aidé à comprendre certaines notions dans le sujet.

Résumé

Cette recherche s'intéresse au rôle de l'intelligence artificielle dans la production des rendus d'architecture, et plus précisément à la manière dont les différents paramètres influencent le style graphique des images générées. L'objectif est d'identifier les éléments les plus déterminants dans le style graphique, tels que le prompt, le CFG, le denoising strength, le SEED ou encore l'image de référence. À travers l'outil Stable Diffusion, cette étude propose une série d'expérimentations permettant d'observer comment chaque paramètre agit sur la composition visuelle, la lumière, les textures ou l'atmosphère générale d'une image. En s'appuyant sur une méthode comparative et des références issues de travaux existants (notamment ceux d'Emma Kaidi sur les styles graphiques), la recherche vise à comprendre jusqu'où l'utilisateur peut contrôler le style final face à l'autonomie de l'algorithme. L'objectif final est de questionner la place de l'auteur dans un processus créatif de plus en plus assisté par l'IA.

Mots clés

Intelligence artificielle, rendus, stable diffusion, prompt, denoising

Abstract

This research explores the role of Artificial Intelligence in the production of architectural renderings, specifically focusing on how various parameters influence the graphic style of generated images. The objective is to identify the most decisive factors in determining graphic style, such as prompts, Classifier Free Guidance (CFG), denoising strength, the SEED, and reference images. Using Stable Diffusion as a primary tool, this study conducts a series of experiments to observe how each parameter affects visual composition, lighting, textures, and the overall atmosphere of an image. Based on a comparative methodology and existing literature (notably the work of Emma Kaidi on graphic styles), this research aims to understand the extent to which a user can control the final aesthetic output versus the algorithm's autonomy. Ultimately, this work questions the author's role within a creative process increasingly assisted by AI.

Table des matières

- 1. INTRODUCTION**
- 2. PROBLEMATIQUE**
- 3. HYPOTHESE**
- 4. OBJECTIF DE LA RECHERCHE**
- 5. GLOSSAIRE**
- 6. ETAT DE L'ART**
 - A. INTRODUCTION A L'INTELLIGENCE ARTIFICIELLE ET AUX RENDUS D'ARCHITECTURE
 - B. MACHINE LEARNING ET DEEP LEARNING
 - C. LES GANS
 - D. STABLE DIFFUSION ET LES MODELES DE DIFFUSION
 - E. LES RENDUS D'ARCHITECTURE ET LES IMAGES DE SYNTHESE
- 7. PRESENTATION DU CORPUS**
 - A. STYLE GRAPHIQUE EMMA KAIDI
 - B. TAXINOMIE DES STYLE GRAPHIQUE
- 8. ETAT DE L'ART DES METHODOLOGIES**
- 9. METHODE DE RECHERCHE**
 - A. CHOIX DE L'OUTIL D'EXPERIMENTATION
 - B. PROCESSUS D'EXPERIMENTATION
 - C. METHODE D'EVALUATION
- 10. RESULTAT DE L'EXPRIMENTATION**
 - A. VARIATION PROMPT
 - B. VARIATION DENOISING
 - C. VARIATION IMAGE DE REFERENCE
 - D. VARIATION CFG
 - E. VARIATION SEED
- 11. CONFRONTATION AVEC LES HYPOTHESES**
- 12. SYNTHESE GLOBALE**
- 13. LIMITES ET PERSPECTIVES**
- 14. CONCLUSION GENERALE**
- 15. REFERENCES**

INTRODUCTION

L'architecte, en tant que créateur de projet, est constamment confronté à la nécessité de convaincre que ce soit un client, un maître d'ouvrage ou un jury. Ce processus de persuasion repose en grande partie sur les images de synthèse, notamment les rendus d'architecture, intérieur et extérieur.

Et dans les rendus d'architecture l'éléments le plus important, plus que le réalisme de l'image, plus que le contenu de l'image, c'est le style graphique qui joue un rôle central. Le style graphique transmet une émotion, une ambiance, une intention. Il raconte une histoire et une sensibilité voulu par l'architecte.

Avec l'émergence de l'intelligence artificielle, notamment des outils de génération d'image, que ce soit par du texte ou par des images de base (des croquis, des volumétries et tout ce qui s'en suit), le processus de création d'une image esthétique et réaliste devient de plus en plus facile, et accessible à monsieur tout le monde.

L'IA propose donc des images partiellement cohérentes et séduisantes, en jouant sur la lumière, les textures et tout ce qui influence directement le style graphique de l'image. Néanmoins, très souvent les images par l'IA semblent correspondre à ce que l'on attend : elles sont esthétiques, séduisantes aux premiers abords, mais elles ne traduisent pas toujours ce que l'on voulait vraiment exprimer à la base, elle donne un sentiment de satisfaction incomplète. En effet, l'outil guide la main de l'utilisateur, oriente le rendu final. On a l'illusion de décider de ce que l'image ressemblera vers la fin, alors qu'en réalité, c'est l'algorithme qui nous propose un certain standard graphique.

Ce constat m'a amené à m'interroger plus précisément sur ce qui influence le style graphique dans les images générées par IA, et qu'est-ce qui oriente le résultat obtenu vers tel ou tel univers visuel.

C'est à partir de cette réflexion que j'ai décidé d'orienter mon travail de recherche sur l'impact des différents paramètres (prompt, image de base, autre élément technique) de génération d'image sur le style graphique des images d'architecture produites par l'IA.

PROBLEMATIQUE

Cette réflexion sur le style graphique des images de synthèses, et mon expérience sur l'utilisation des outils de génération d'image par de l'IA, que ce soit par du texte ou par des images de référence, m'as poussé à me poser donc la problématique suivante :

Comment les différents paramètres et formulations de prompts influencent-ils les résultats produits dans les processus de l'Image to-Image assistés par intelligence artificielle ? et quels sont les paramètres les plus déterminants pour le style graphique de l'image finale ?

HYPOTHESE

- Hypothèse 1 : La précision lexicale dans la formulation des prompts influences directement le style graphique de l'image générée.
- Hypothèse 2 : est-ce que l'ordre des mots exerce une importance sur le style graphique.

OBJECTIF DE LA RECHERCHE

L'objectif principal de cette recherche est de comprendre comment les différents paramètres de l'IA influencent le style graphique des images de rendus d'architecture généré par l'intelligence artificielle. Il s'agit de déterminer dans quelle mesure l'utilisateur peut orienter le résultat final, et comment les éléments comme le prompt (ou autre paramètres) participent à la production d'un style graphique particulier. À travers cette exploration, le but est d'identifier la place de l'utilisateur dans un processus de plus en plus assisté.

GLOSSAIRE

L'image de synthèse :

Une image de synthèse est une représentation numérique d'une scène ou d'un objet, créée à l'aide d'un ordinateur et de logiciels spécialisés, avec l'objectif de visualiser des concepts ou des réalités de manière réaliste ou stylisée

Rendu :

Processus de génération d'une image à partir d'un modèle 3D, en simulant l'éclairage, les textures et les matériaux. Les moteurs de rendu sont utilisés pour créer des images de synthèse de haute qualité

Texturing :

"Le texturing consiste à définir l'aspect d'un objet 3D. Il définit le réglage des textures donc la manière dont celles-ci vont répondre à la lumière (réflectivité, transparence, couleurs, relief...)"

[KAIDI, 2021]

GANs (Réseaux Antagonistes Génératifs) :

Modèles d'apprentissage profond composés de deux réseaux neuronaux, un générateur et un discriminateur, qui s'entraînent en compétition. Le générateur crée de nouvelles données (par exemple des images), tandis que le discriminateur évalue si les données sont réelles ou générées.

[MARSULT et NGUYEN, 2022]

Latent space (espace latent)

Représentation vectorielle abstraite des caractéristiques d'un ensemble de données, appris par un modèle d'apprentissage automatique, tel qu'un GAN. L'espace latent permet de manipuler les caractéristiques des données, et donc de modifier les images produites par un GAN, par exemple

"We ... have access to an efficient, low-dimensional latent space in which high-frequency, imperceptible details are abstracted away. Compared to the high-dimensional pixel space, this space is more suitable

for likelihood-based generative models..."

[ROMBACH ET AL., 2022] HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS. 2022.

Autoencodeur / VAE (encodeur-décodeur)

Couple encodeur (E) → compresse l'image en latent z ; décodeur (D) → reconstruit l'image. Dans LDM, le VAE assure une compression perceptuelle puis un décodage vers l'espace image.

"We train an autoencoder which provides a lower-dimensional representational space perceptually equivalent to the data space... The encoder E encodes x into a latent representation $z = E(x)$, and the decoder D reconstructs the image from the latent, giving $\tilde{x} = D(z)$."

[ROMBACH ET AL., 2022] HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS. 2022.

U-Net (réseau de débruitage)

Architecture conv-deconv avec skip connections servant de backbone temporel pour prédire et retirer le bruit à chaque étape.

"The neural backbone ... is realized as a time-conditional UNet [71]."

[ROMBACH ET AL., 2022] HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS. 2022.

Image-to-Image / Inpainting

Conditionnement par image/masque : image-to-image transfère structure/style depuis une référence ; inpainting remplit des zones masquées en cohérence avec le contexte.

"Inpainting is the task of filling masked regions of an image with new content ... We evaluate how our general approach for conditional image generation compares to specialized ... inpainting."

Prompt

Instruction textuelle qui guide la génération : elle transmet l'intention (contenu, style, ambiance, matériaux, composition) au modèle. Des prompts bien conçus augmentent l'alignement au cahier des charges et la créativité.

"Prompts are not merely simple words or phrases; they act as communication bridges between designers and machine learning algorithms. Well-designed Prompts can accurately convey design intent and requirements..."

[ZHANG ET LIU, 2024] XIN ZHANG; WENWEN LIU. BOOSTING ARCHITECTURAL GENERATION VIA PROMPTS: REPORT. 2024.

Negative Prompt

Variante de la guidance : on fournit un prompt négatif pour repousser des attributs indésirables pendant le débruitage (extension de C.F.G. au texte négatif).

"In classifier-free guidance, the predicted noise ϵ at each sampling step is ... extrapolated towards the conditional prediction and pushed away from the unconditional one... by replacing the empty prompt with a 'negative' prompt P_n , we can push away ..."

[TUMANYAN ET AL., 2022] MORE CONTROL FOR FREE! IMAGE SYNTHESIS WITH SEMANTIC DIFFUSION GUIDANCE. 2022.

CFG (Classifier-Free Guidance scale)

Échelle de guidage qui renforce l'alignement au texte : plus la valeur est haute, plus l'image suit le prompt (souvent au prix d'une créativité plus faible).

"Applying classifier-free diffusion guidance greatly boosts sample quality..."

Denosing strength

Processus central : le modèle retire progressivement le bruit pour reconstruire l'image. En image-to-image, le *denoising strength* règle le degré de réinterprétation : faible → proche de la source ; élevé → plus créatif / moins fidèle.

"Diffusion models ... learn a data distribution by gradually denoising a normally distributed variable, which corresponds to learning the reverse process of a fixed Markov Chain of length T ."

[ROMBACH ET AL., 2022] ROBIN ROMBACH, ANDREAS BLATTMANN, DOMINIK LORENZ, PATRICK ESSER, BJÖRN OMMER. HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS. 2022.

Seed (graine)

Valeur aléatoire qui fixe l'initialisation du bruit : à paramètres identiques, même seed = même image, changer le seed varie la composition/détails sans changer le prompt.

"The seed can have a large effect when using certain prompts... we want to describe the image induced by a single seed as specifically as possible."

[DECKERS ET AL., 2024] MANIPULATING EMBEDDINGS OF STABLE DIFFUSION PROMPTS. 2024.

Sampler / Scheduling (procédure d'échantillonnage)

Le *sampler* (DDIM, etc.) définit comment le débruitage progresse en N étapes ; le *scheduling* règle le rythme de suppression du bruit. Ensemble, ils gouvernent le compromis qualité ↔ vitesse.

"We ... compare models ... for different numbers of denoising steps with the DDIM sampler and plot it against FID-scores."

[ROMBACH ET AL., 2022] HIGH-RESOLUTION IMAGE SYNTHESIS WITH LATENT DIFFUSION MODELS. 2022.

ETAT DE L'ART

Introduction à l'Intelligence Artificielle et aux rendus d'architecture

L'intelligence artificielle (IA) et les rendus d'architecture sont deux éléments qui peuvent se combiner pour améliorer la façon dont on présente un projet en architecture. Les rendus doivent être une représentation fidèle d'un projet, L'IA dans ce cadre permet de générer des visuels réalistes en automatisant des tâches complexes.

Dans le cadre l'intelligence artificiel il y a Les générateurs d'images text-to-image qui facilitent la création rapide de d'images à partir de descriptions textuelles (prompt), ce qui permet aux architectes de s'inspirer de plusieurs images générées. Et il y a aussi des améliorateurs d'image (image-to-image) ou l'IA vient apporter une amélioration visuelle de l'image de synthèse ou d'un dessin/croquis et dans ce cas la modélisation 3D reste une étape fondamentale pour garantir la cohérence et le réalisme des rendus.

Machine learning et deep learning

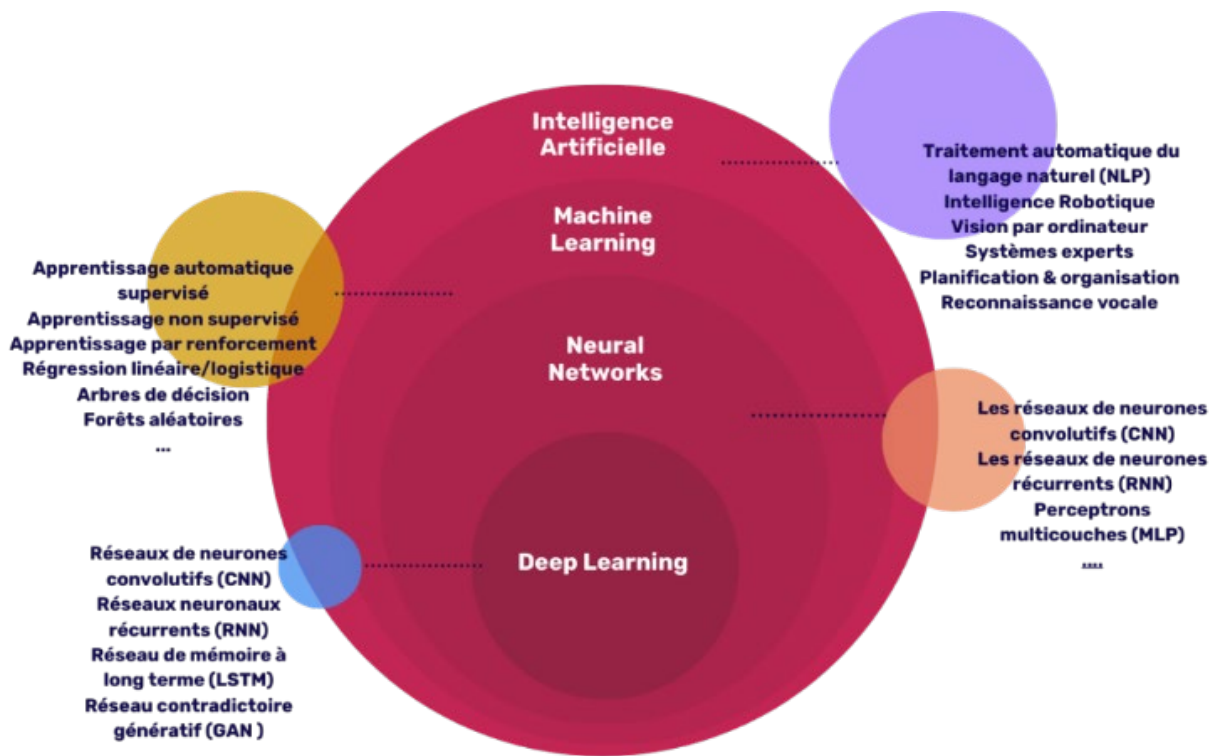


FIGURE 1 : LES DIFFERENTS DOMAINES ET SPECIALISATIONS DE L'INTELLIGENCE ARTIFICIELLE

SOURCE: [HTTPS://INVENTIV-IT.FR/AI-MACHINE-LEARNING-DEEP-LEARNING/](https://inventiv-it.fr/ai-machine-learning-deep-learning/)

L'apprentissage automatique, ou machine learning, constitue l'un des piliers de l'intelligence artificielle et c'est aussi parmi les premiers éléments de l'IA. Il offre aux machines la capacité d'apprendre et de s'améliorer grâce à des données. Et du machine learning provient le deep learning qui utilise des réseaux de neurones profonds et qui est le sous domaine de l'ia spécifique a la génération d'image et l'amélioration d'image, ce sont ces modèles qui sont utilisé dans ce cadre, ils sont capables de traiter d'importantes quantités de données. Ces systèmes permettent de créer des représentations complexes, adaptées à des tâches comme la reconnaissance d'images, la synthèse ou encore la traduction automatique.

[BEYAN et ROSSY, 2023]

Parmi ces modèles, les réseaux antagonistes génératifs (GANs) qui jouent un rôle clé dans ce domaine. Son fonctionnement est une compétitive entre deux réseaux de neurones : un générateur qui crée des images et un discriminateur qui évalue leur authenticité. Cette interaction se répète et le plus elle se répète le plus le générateur est capable de générer des images réalistes, et donc le modèle devient le plus performant possible.

[ZHOU et PARK, 2021]. [BEYAN et ROSSY, 2023]

Il y a aussi un point important dans ces modèles c'est que L'architecture des réseaux neuronaux influence leur performance et leur spécialisation par exemple dans les modèles de Deep Learning les plus utilisés (les CNN et les RNN) chacun son rôle : Les réseaux convolutifs (CNN) se concentrent sur l'analyse et la reconnaissance d'images, tandis que les réseaux récurrents (RNN), spécialisés dans les séquences, traitent des données telles que les vidéos ou les textes. Ces variations structurelles sont adaptées à des besoins spécifiques dans le domaine de la génération d'images et de l'amélioration d'image.

Dans le domaine des IA il existe aussi un concept important à comprendre spécifiquement dans le machine Learning, et c'est le processus d'apprentissage, c'est la méthode d'apprentissage du modèle, L'intelligence artificielle repose sur trois processus d'apprentissage : supervisé, non supervisé et par renforcement. Les deux premiers sont basés sur des données créées par l'humain par exemple des images étiquetées (on donne une image de banane et on dit que c'est une banane) et pour le non supervisé on ne dit pas que c'est une banane et on laisse l'algorithme catégoriser le tout. Quand à l'apprentissage par renforcement, basé sur des interactions avec l'environnement via des récompenses ou des pénalités (interaction avec un jeu d'échec par exemple), il est particulièrement utile pour optimiser des processus complexes en architecture.

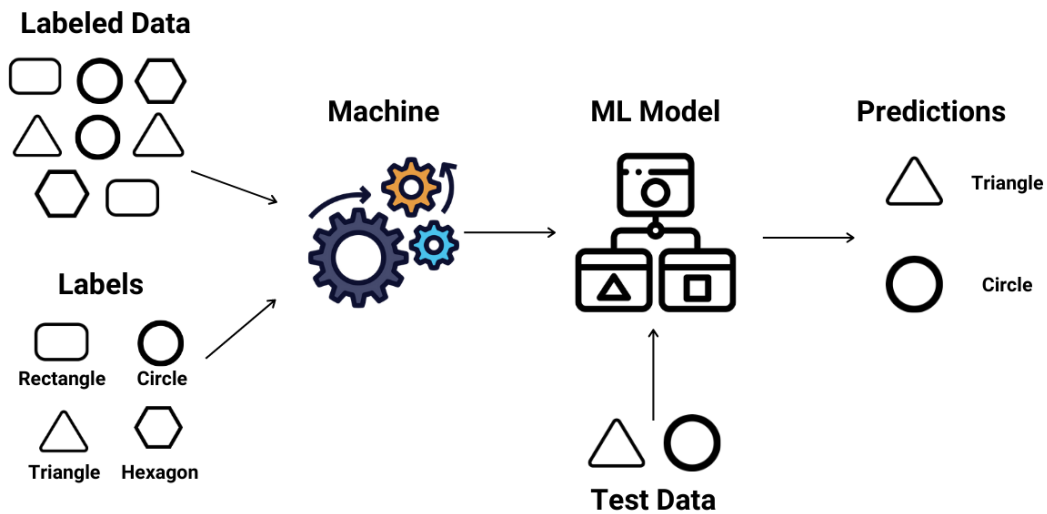


FIGURE 2 : SCHEMA EXPLIQUANT L'APPRENTISSAGE NON-SUPERVISE
 SOURCE : [HTTPS://BLENT.AI/BLOG/A/APPRENTISSAGE-SUPERVISE-DEFINITION](https://blent.ai/blog/a/apprentissage-supervise-definition)

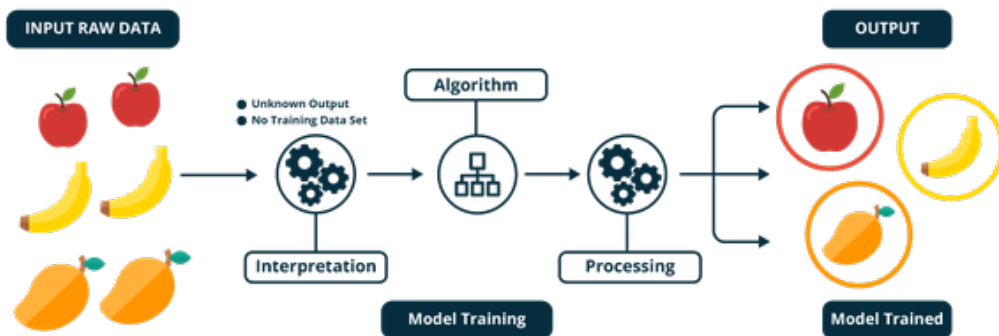


FIGURE 3 : SCHEMA EXPLIQUANT L'APPRENTISSAGE NON-SUPERVISE
 SOURCE : [HTTPS://BLENT.AI/BLOG/A/APPRENTISSAGE-SUPERVISE-DEFINITION](https://blent.ai/blog/a/apprentissage-supervise-definition)

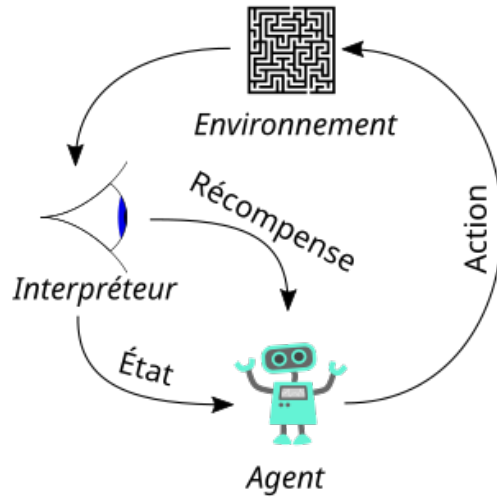


FIGURE 4 : SCHEMA EXPLIQUANT L'APPRENTISSAGE NON-SUPERVISE

SOURCE : <https://blent.ai/blog/a/apprentissage-supervise-definition>

Cependant, la performance des modèles dépend de la qualité et de la diversité des données d'entraînement, qui réduisent les biais et améliorent leur capacité à s'adapter. Un équilibre est nécessaire entre la complexité des modèles et le volume de données disponibles pour éviter le surapprentissage.

[BEYAN et ROSSY, 2023]

Les GANs

Introduction historique : Les GANs (Réseaux antagonistes génératifs), sont des modèles de Deep Learning apparus en 2014, ils ont connu une véritable révolution depuis leur création. Leur force réside dans leur capacité à saisir et à mémoriser des multitudes de données, souvent hétérogènes, tout en s'ajustant à la structure du monde qui nous entoure. Ce phénomène les rend particulièrement puissants dans des domaines créatifs comme l'architecture.

Depuis leurs débuts, les GANs ont considérablement évolué. Leurs performances ont été grandement améliorées, notamment en matière de stabilité, de convergence, de gestion de la mémoire, et de la qualité des images générées. Plusieurs versions secondaires de ces réseaux sont apparues après, comme DCGAN, W-GAN, BEGAN, PG-GAN, SD-GAN, et StyleGAN2, et chacun sa particularité et son utilité

[MARSAULT et NGUYEN, 2022]

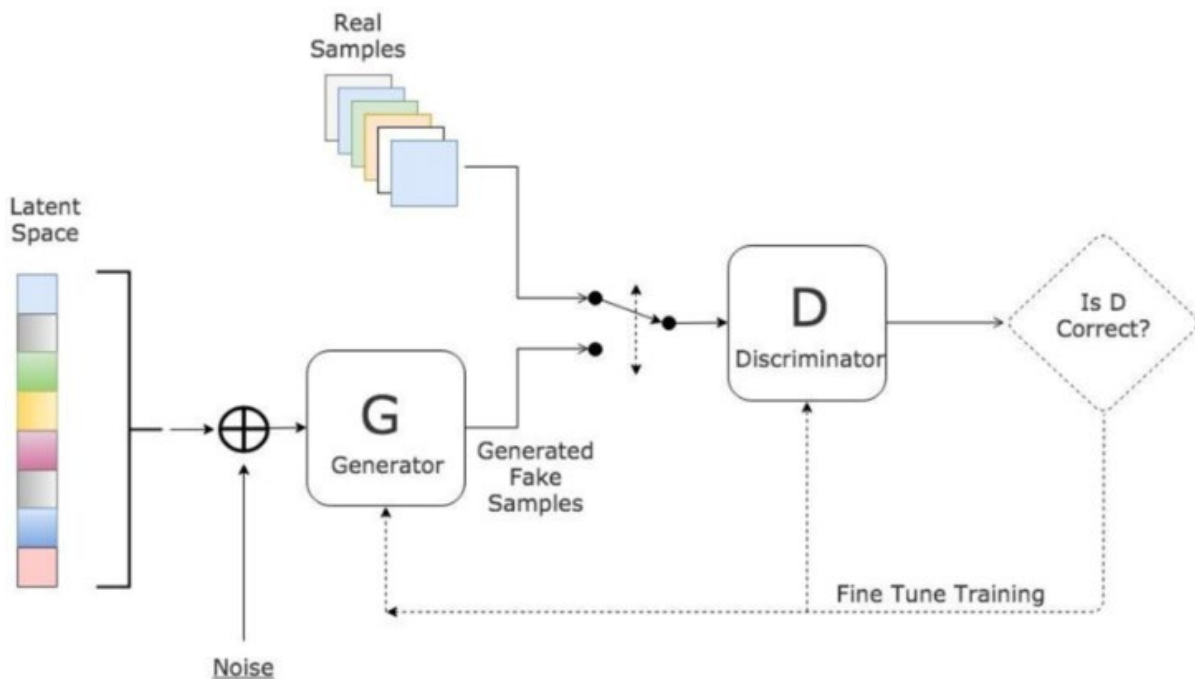


FIGURE 5 : SCHEMA CONCEPTUEL DES GANs

SOURCE : [HTTPS://WWW.RESEARCHGATE.NET/FIGURE/GENERAL-SCHEMA-OF-GAN-ARCHITECTURE-14_fig2_36111921](https://www.researchgate.net/figure/GENERAL-SCHEMA-OF-GAN-ARCHITECTURE-14_fig2_36111921)

Le fonctionnement d'un GAN repose sur une compétition entre deux modèles d'intelligence artificielle qui s'entraînent mutuellement :

- Le Générateur (G) : Son rôle est de créer des données factices. Il prend en entrée des vecteurs provenant d'un espace latent mélangés à du Noise (bruit) pour générer des échantillons d'image. Son but est de devenir si performant que ses créations ressemblent aux données réelles.
- Le Discriminateur (D) : Il agit comme un juge. Il reçoit alternativement des images réelles depuis les données et les images produits par le générateur. Il compare à chaque fois les deux images (donnée/image du générateur) et Son rôle est de déterminer si ce qu'il voit est "vrai" ou "faux".

A la fin le générateur sera capable de générer des échantillons qui seront très fidèle et très réaliste avec les images réelles.

Stages d'apprentissage du GANS

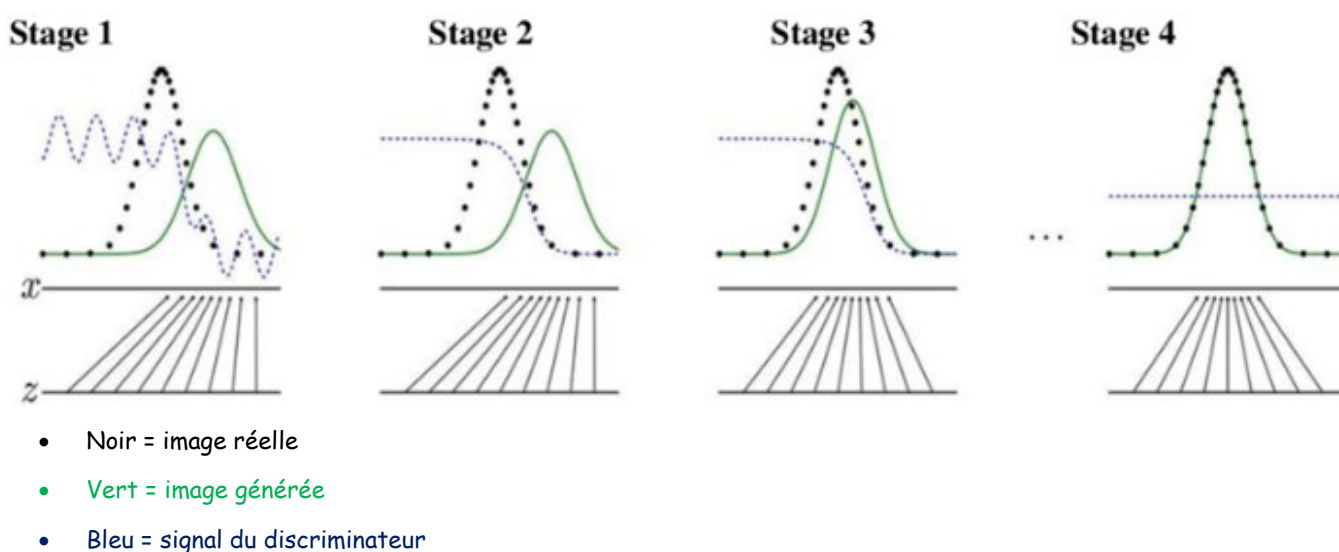


FIGURE 6 : ILLUSTRATION DE L'APPRENTISSAGE PROGRESSIF D'UN RESEAU ANTAGONISTE GENERATIF (GAN) ET CONVERGENCE DES DISTRIBUTIONS

SOURCE: [HTTPS://WWW.RESEARCHGATE.NET/FIGURE/GRADUAL-LEARNING-OF-GAN-BLACK-TRUE-GREEN-GENERATED-BLUE-DISCRIMINATOR-ERROR-SIGNALS-5_FIG3_361111921](https://www.researchgate.net/figure/GRADUAL-LEARNING-OF-GAN-BLACK-TRUE-GREEN-GENERATED-BLUE-DISCRIMINATOR-ERROR-SIGNALS-5_FIG3_361111921)

L'image montre les stages d'apprentissage des GANs en expliquant la compétition entre le Générateur (G) et le Discriminateur (D) pour atteindre un Équilibre de Nash.

Les courbes de probabilité représentent : la distribution réelle (Noire), cible ultime de l'apprentissage ; la distribution générée par G (Verte), qui cherche à imiter la distribution réelle ; et le signal d'erreur du Discriminateur (Bleu), représentant la confiance de D quant à l'authenticité des échantillons.

1. Stage 1 (Faible Convergence) : Le Générateur est initialement faible, et sa distribution (Verte) est éloignée de la cible (Noire). Le Discriminateur est très efficace et confiant, produisant un signal d'erreur fort et variable (Bleu) là où les distributions se chevauchent à peine. Ce gradient puissant guide G pour ajuster ses paramètres.
2. Stage 2 (Amélioration) : Sous l'effet du gradient, G commence à structurer sa distribution (Verte) pour se rapprocher de la distribution réelle. La tâche de D devient plus difficile, et son signal d'erreur (Bleu) s'atténue dans la zone de chevauchement, traduisant une incertitude croissante.
3. Stage 3 (Convergence Avancée) : Les deux distributions (Noire et Verte) sont désormais largement superposées. Le Discriminateur (D) atteint un état d'incertitude maximale et n'arrive plus à distinguer les échantillons réels des échantillons générés. L'erreur (Bleu) devient quasi uniforme et tend vers une probabilité de 50% (0.5), prouvant que G est devenu très performant.
4. Stage 4 (Équilibre Final) : L'apprentissage converge vers un Équilibre de Nash. La distribution générée (Verte) a réussi à imiter et à recouvrir presque parfaitement la distribution réelle (Noire). Le Discriminateur est complètement trompé, ce qui signifie que le Générateur a réussi à reproduire les caractéristiques statistiques des données d'entraînement.

En architecture, les GANs sont utilisés pour générer des plans, des façades, et des perspectives urbaines. Ils permettent aux architectes de repousser les frontières de leur imagination numérique, en combinant des

détails à différentes échelles et styles. L'un des grands atouts des GANs réside dans leur capacité à aborder des problématiques complexes et subjectives, tout en évitant l'obligation d'explication détaillée.

Une autre avancée clé dans l'utilisation des GANs est l'exploration de leur espace latent. En manipulant cet espace il devient possible de jouer avec les caractéristiques des images générées, d'affiner certains détails, ou même d'y projeter des données spécifiques fournies par l'utilisateur. Cependant, l'interprétation sémantique de cet espace latent reste un défi. Manipuler les données générées par les GANs, tout en en comprenant profondément le sens et l'impact, n'est pas encore une tâche simple

[MARSAULT et NGUYEN, 2022]

Stable diffusion et Les modèles de diffusion

Stable Diffusion est un modèle de génération d'images qui repose sur les modèles de diffusion (diffusion model), ces derniers sont une famille d'algorithmes qui génèrent une image en apprenant à inverser un processus de dégradation progressive de l'image. Et cela s'effectue en deux principales étapes :

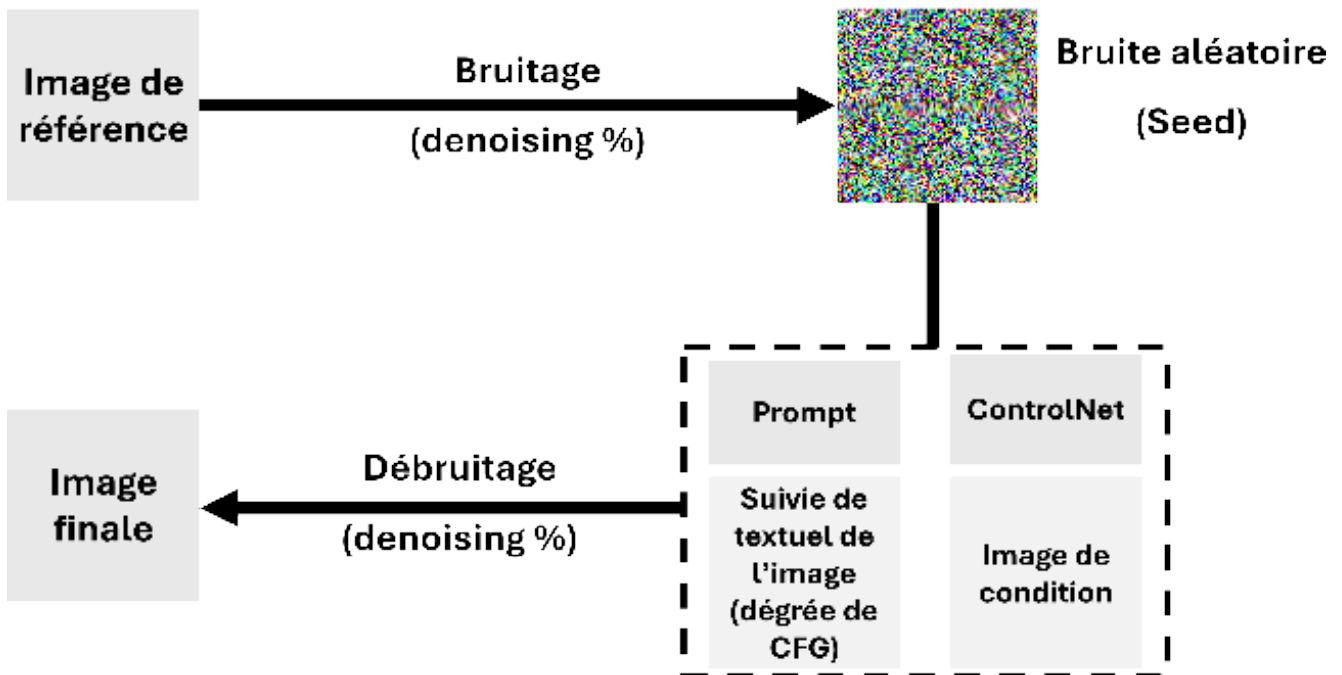


FIGURE 7 : SCHEMA EXPLICATIF SIMPLIFIE DU FONCTIONNEMENT DU MODELE DE DIFFUSION UTILISEE DANS STABLE DIFFUSION
SOURCE : SCHEMA PERSONNEL

1. **Le processus de diffusion (Forward Process)** : L'image réelle est progressivement dégradée en lui ajoutant du bruit gaussien étape après étape, jusqu'à devenir un bruit pur.
2. **Le processus inverse (Reverse Process)** : le modèle apprend à retirer le bruit étape par étape, en prédisant le bruit ajouté à chaque étape. Cette prédiction est effectuée par le UNet conditionné par le texte.

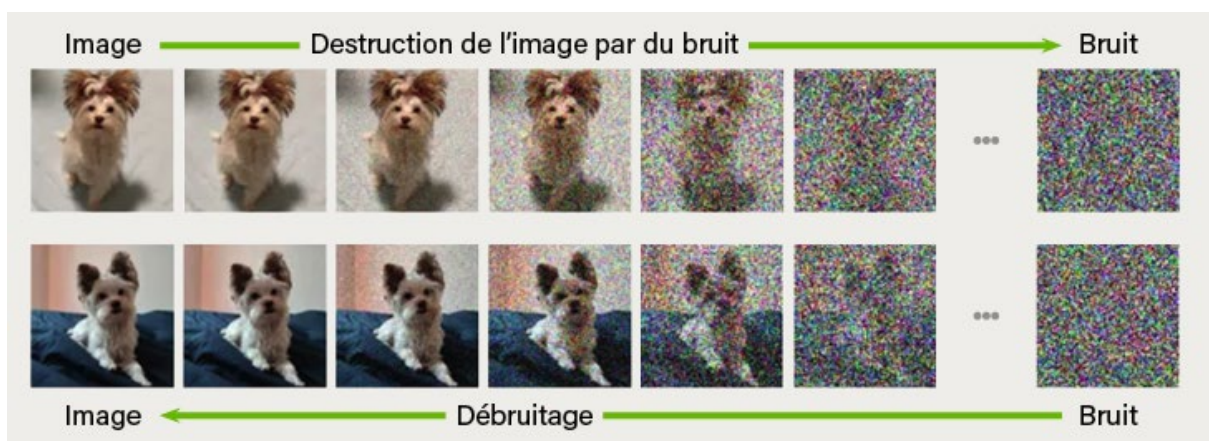


FIGURE 8 : EXEMPLE QUI MONTRE LE PROCESSUS DES MODELES DE DIFFUSION
SOURCE : [HTTPS://WWW.POURLASCIENCE.FR/SR/LOGIQUE-CALCUL/DU-TEXTE-A-L-IMAGE-CAP-FRANCHI-POUR-L-IA-26135.PHP](https://www.pourlascience.fr/sr/logique-calcul/du-texte-a-l-image-cap-franchi-pour-l-ia-26135.php)

Le trilemme génératif

C'est un concept important à comprendre dans l'étude des modèles d'IA générative. Un modèle ne peut généralement pas optimiser simultanément les trois dimensions suivantes :

- La qualité des échantillons produits (High Quality Samples)

- La rapidité d'échantillonnage (Fast Sampling)
- La diversité et la couverture des modes (Mode Coverage / Diversity)

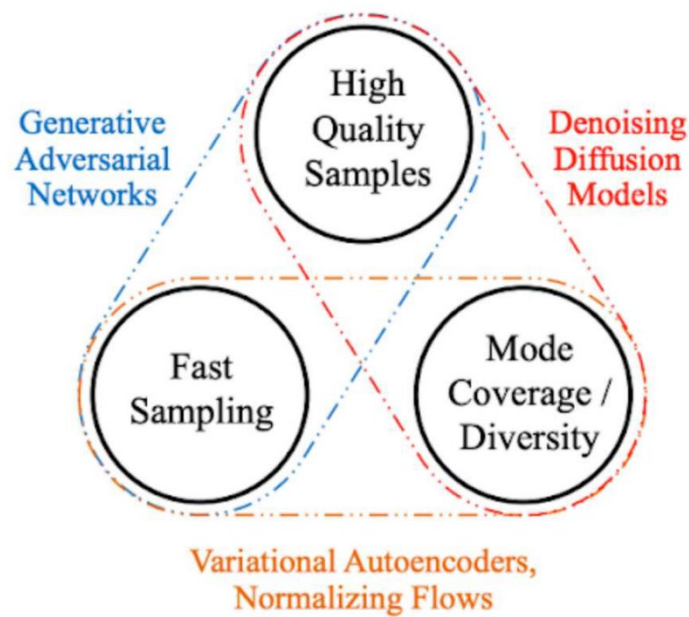


Figure 1. Generative learning trilemma

FIGURE 9 : LE TRILEMME DE L'APPRENTISSAGE GENERATIF
SOURCE : SCIENCE ETONNANTE

Chaque famille de modèles excelle dans un ou deux de ces axes au détriment du troisième par exemple, les GANs privilégient la qualité et la rapidité (manque de diversité), les VAEs et flows quant à eux favorisent la diversité tout en sacrifiant la qualité, et finalement les modèles de diffusion, comme Stable Diffusion, maximisent la qualité et la diversité, mais sont plus lents.

Spécificité des modèles de diffusion

La grande innovation de Stable Diffusion est de réaliser le processus de débruitage dans un espace latent compressé, obtenu via un VAE (Variational Autoencoder).

L'article montre une image "bruitée" typique en 64×64 pixels, upscalée ensuite. Cela permet une rapidité d'exécution du modèle (fois 8)

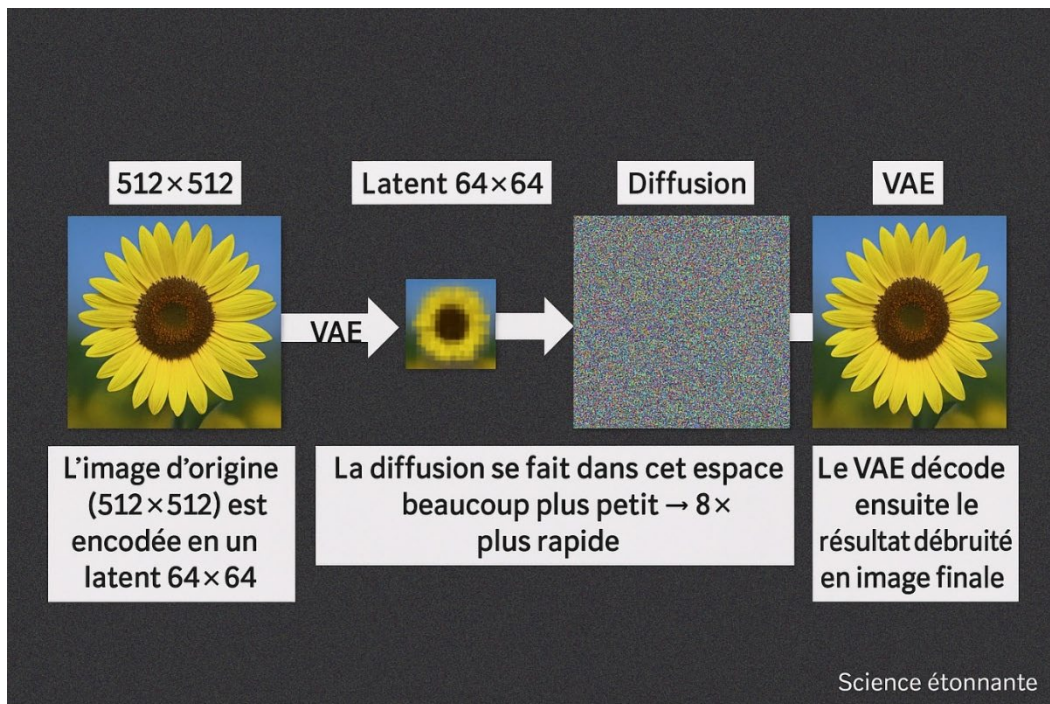


FIGURE 10 : IMAGE EXPLICATIVE DE L'ESPACE LATENT (COMPRESSION DE L'IMAGE)
SOURCE : SCIENCE ETONNANTE

Etapes d'exécution du modèle :

- L'image d'origine (512×512) est encodée en un latent 64×64.
- La diffusion se fait dans cet espace beaucoup plus petit
- Le VAE décode ensuite le résultat débruité en image finale.

Architecture de Stable Diffusion

Stable Diffusion repose sur trois modules principaux :

Le VAE : Encode l'image dans un espace latent compressé, Puis décode le latent final en image 512×512 ou 1024×1024.

L'encodeur texte CLIP : Utilisé pour transformer un **prompt textuel** en vecteur sémantique.

Le UNet de diffusion : Le U-Net un réseau en forme de « U » composé d'une phase d'encodage et d'une phase de décodage.

Dans la partie descendante, le réseau réduit progressivement la résolution grâce à des convolutions et du *max pooling*, ce qui lui permet d'extraire des caractéristiques globales.

Dans la partie ascendante, il reconstruit une représentation plus détaillée de l'image grâce aux *up-convolutions*. Les connexions skip relient les deux parties et permettent de conserver des détails importants.

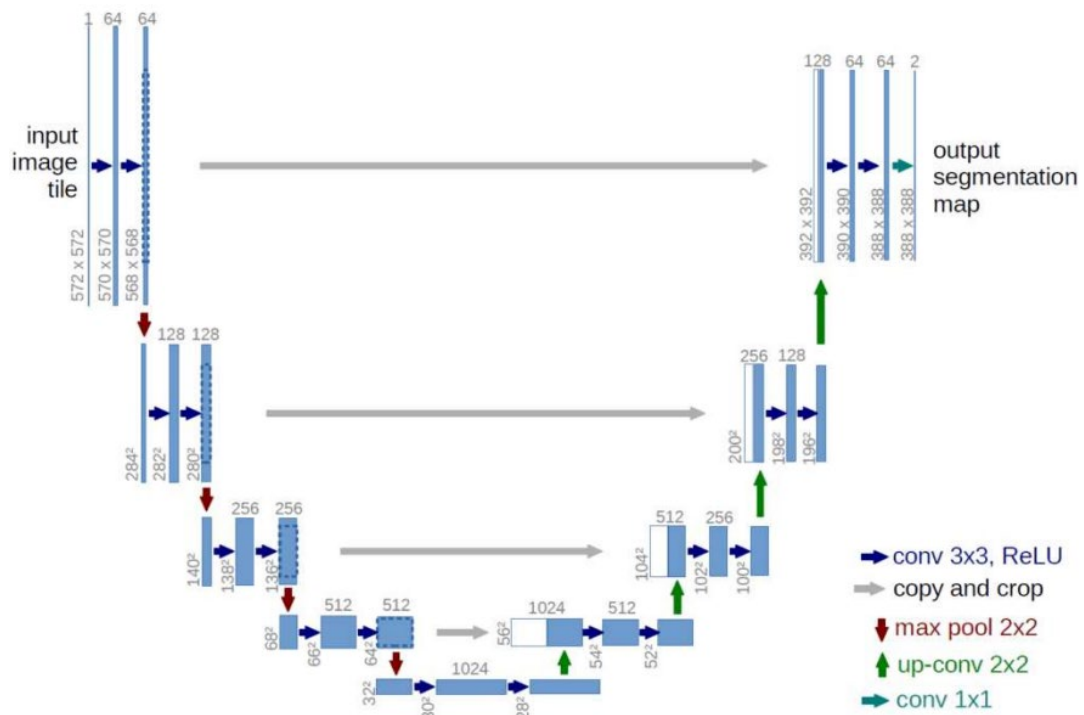


FIGURE 11 : ARCHITECTURE DU UNET
SOURCE : SCIENCE ETONNANTE

Exemple d'utilisation de stable diffusion

Stable Diffusion est utilisée dans plusieurs domaines de création et de retouche d'images, que ce soit pour la suppression ou l'ajout d'éléments, la transformation d'une esquisse vers une image réelle, ou la création d'une image réelle à partir d'un prompt. Par exemple :

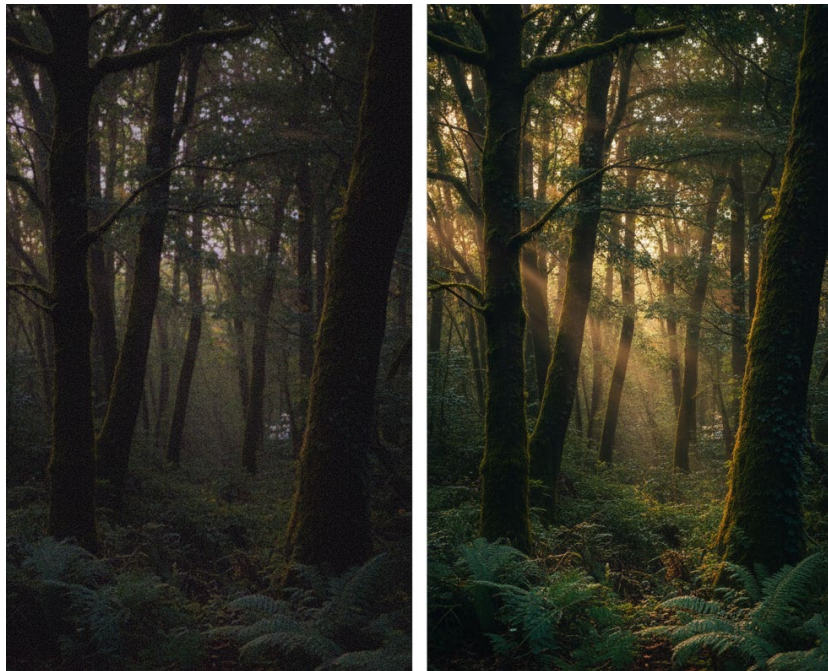


FIGURE 12 : CHANGEMENT DE GRAPHISME, AJOUT DE LUMIERE
SOURCE : IMAGE PERSONNELLE



FIGURE 13 : SUPPRESSION D'ELEMENTS
SOURCE : SCIENCE ETONNANTE



FIGURE 14 : ESQUISSE EN UNE IMAGE REALISTE
SOURCE : SCIENCE ETONNANTE

ControlNet (contrôle structurel)

ControlNet ajoute un réseau parallèle au UNet qui impose une **structure fixe** tout en laissant Stable Diffusion remplir le contenu. Il est donc un élément très important surtout quand on veut garder la même structure globale de l'image. Par exemple : les poses d'un personnage, les dimensions en architecture, etc. ControlNet a donc plusieurs modèles, chaque modèle est spécifique à un besoin précis par exemple :

- **Canny** : pour suivre les contours précis d'un objet.
- **Depth** : pour respecter les distances et le relief (la profondeur).

- **OpenPose** : pour copier exactement la position d'un corps humain.
- **MLSD** : pour analyser les lignes droites (très utile pour l'architecture).
- **SoftEdge** : pour des contours plus souples et moins rigides que Canny.
- **Scribble** : pour transformer un simple gribouillage en image finale.
- **Segmentation** : pour séparer et identifier les différents éléments d'une scène.

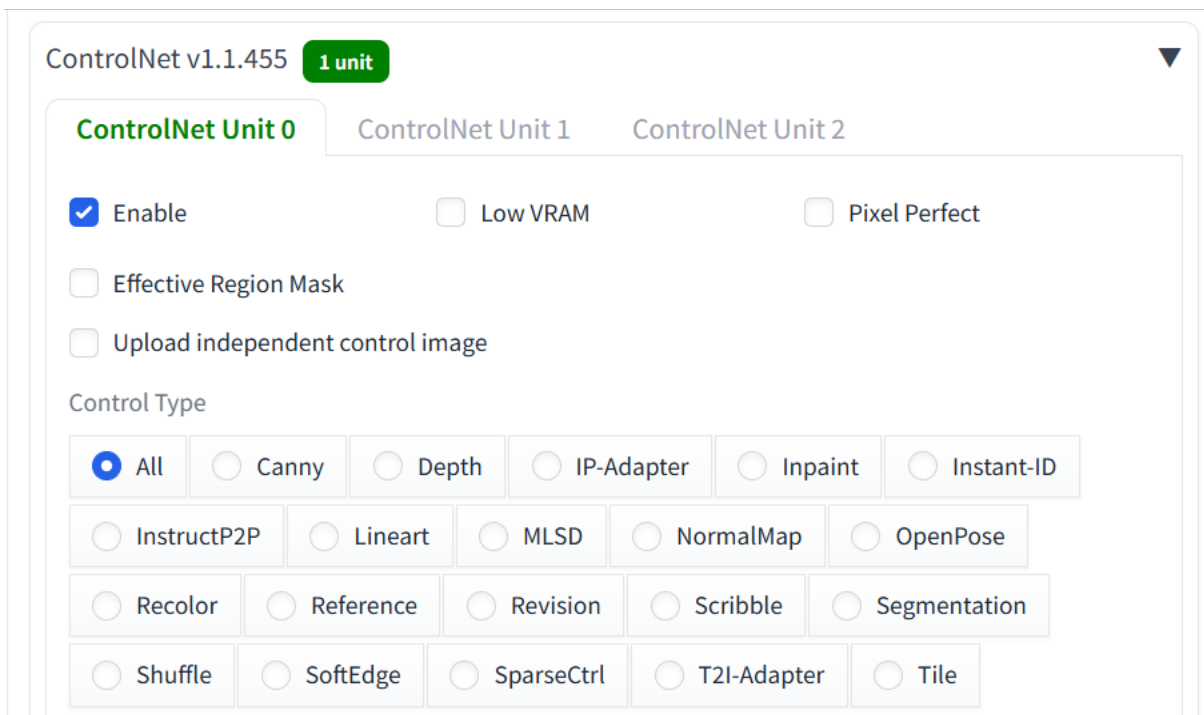


FIGURE 15 : MODELE DE STABLES DIFFUSION
SOURCE : CAPTURE DE STABLE DIFFUSION (CONTROLNET)

Paramètres essentiels de Stable Diffusion

Les paramètres les plus importants sur stable diffusion sont

- **Prompt**
- **Negative prompt**
- **Guidance Scale (CFG)**
- **Denoising strenght**
- **Sampling Steps**
- **Seed**
- **ControlNet**

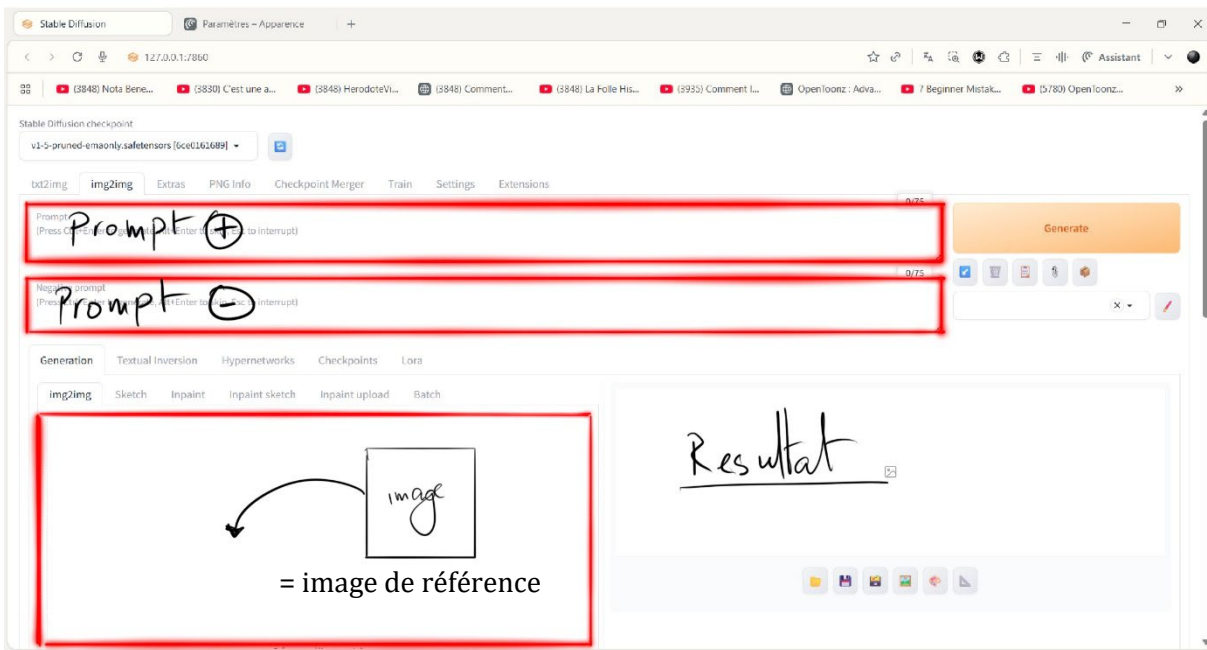


FIGURE 16 : PARAMETRES DE STABLE DIFFUSION
SOURCE : CAPTURE DE L'INTERFACE DE STABLE DIFFUSION

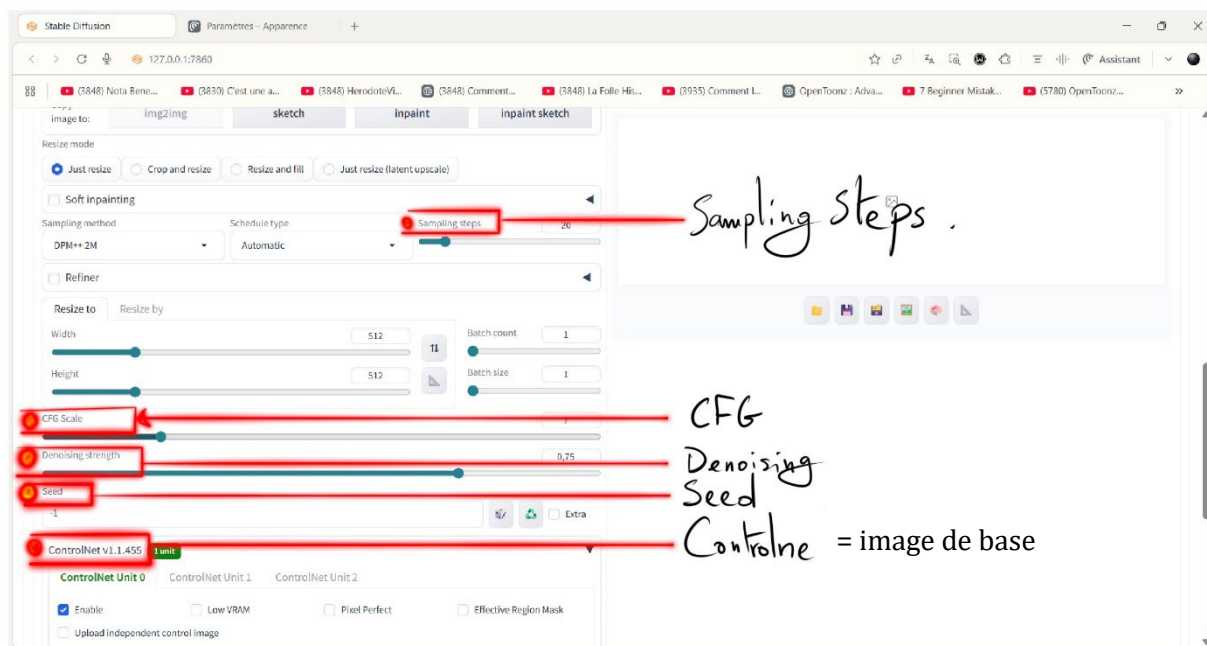


FIGURE 17 : PARAMETRES DE STABLE DIFFUSION
SOURCE : CAPTURE DE L'INTERFACE DE STABLE DIFFUSION

Les Rendus d'architecture et Les images de synthèse

Dans les concours d'architecture, les images de synthèse occupent une place essentielle en tant qu'outils stratégiques de communication, capables de mettre en valeur un projet et ses intentions architecturales pour capter l'attention des jurys. Chaque architecte dispose d'une grande liberté dans leur utilisation, choisissant les représentations qui reflètent le mieux les qualités de son concept. Cette liberté, associée à la nature compétitive des concours, elle incite à perfectionner les rendus pour séduire et se démarquer. Ces images ne se contentent pas d'expliquer le projet, elles mobilisent diverses techniques graphiques pour magnifier la lumière, les matériaux, les volumes et leur relation au contexte, à travers des plans, coupes, élévations et perspectives, chacun jouant un rôle précis. Cependant, aucune solution architecturale n'est universelle : chaque projet est une réponse unique à un cahier des charges, exprimant les choix et

sensibilités de son auteur. Ainsi, les concours deviennent des espaces d'expression où s'affirment identité et créativité, tout en reflétant les avancées techniques et stylistiques de leur époque. [KAIDI, 2021]

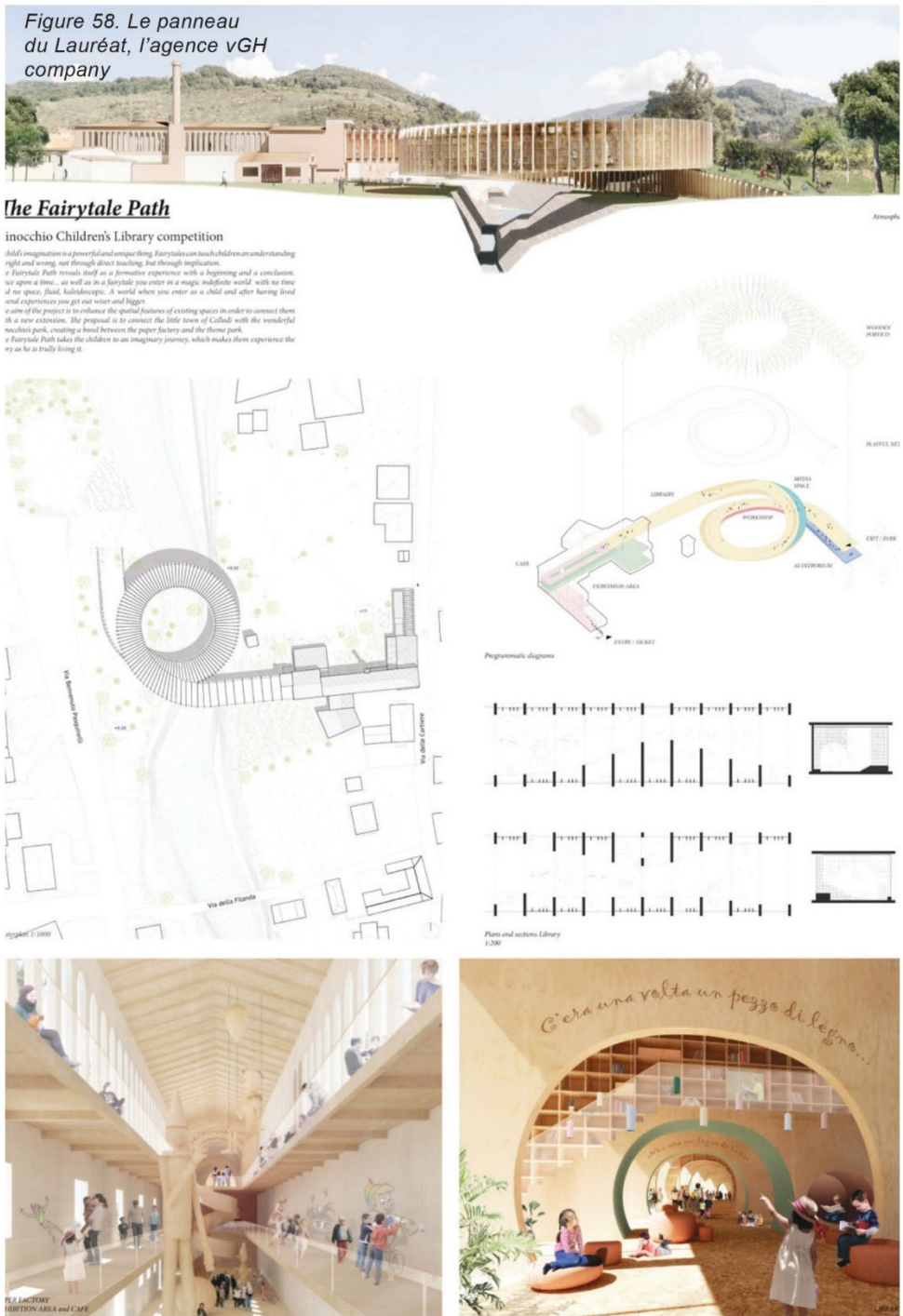


FIGURE 18 : EXEMPLE DE CONCOURS CONTENANT DES IMAGES DE SYNTHÈSES (AGENCE VGH COMPANY)

SOURCE : [KAIDI, 2021]

Les rendus de concours ne se limitent pas à présenter un projet : ils constituent un support d'étude fascinant pour comprendre les styles et les techniques visuelles en vogue. À travers ces rendus, les architectes traduisent leurs intentions avec un usage maîtrisé des outils graphiques, des plans aux perspectives.

[KUPPENS, 2020] et [KAIDI, 2021]

Le style graphique :

« Le style est un moyen de retranscrire un univers, un récit, des idées et une atmosphère. Aucun style ne semble être la clé du succès d'un concours, mais plutôt un moyen de faire passer subtilement des idées et des messages sensibles. [...] Le style des images de synthèse pourrait incarner la signature architecturale de son auteur, comme la patte graphique d'un dessinateur. » [KAIDI, 2021]

Autrement dit le style graphique n'est pas seulement une esthétique, il traduit aussi l'intention de l'architecte et devient sa marque de fabrique.

Les critères d'évaluation des images de synthèse

Les rendus architecturaux s'expriment à travers une palette variée de styles, qui vont du photoréalisme saisissant à des approches plus conceptuelles ou abstraites.

Dans ce contexte, la démarche d'Emma Kaidi (2021) s'avère particulièrement structurante : elle opte pour une sélection d'images issues de concours et met en place des critères graphiques objectifs – lumière, composition, textures – qui sont ensuite traduits en grilles d'analyse et en tableaux de classification. Cette méthode, basée sur l'observation systématique, permet d'appréhender finement la diversité des styles graphiques dans l'architecture contemporaine.

Le photoréalisme, s'impose comme la norme pour transmettre une vision proche de la réalité. Cependant, il reste une multitude d'options pour illustrer un projet. [KAIDI, 2021]. En revanche Les architectes ne se limitent pas à un seul style de représentation. Ils explorent et mélangent différents types de rendus, créant des combinaisons pour renforcer l'impact visuel et argumentatif de leurs propositions. Cette liberté graphique permet d'ajuster le message en fonction des objectifs du projet. [KUPPENS, 2020]. D'un point de vue stylistique, les images peuvent être regroupées en deux grandes catégories : celles qui misent sur un rendu réaliste, cherchant à reproduire fidèlement l'apparence du projet dans son futur contexte, et celles qui s'en éloignent, privilégiant une lecture plus artistique ou suggestive. Même dans le réalisme, il existe des degrés subtils de traitement qui différencient les styles. [KAIDI, 2021]

Ainsi, il y a Plusieurs critères pour évaluer la qualité des rendus architecturaux :

- Le réalisme, qui peut aller du strict mimétisme à des interprétations plus libres.
- La composition de la scène, qui structure la lecture visuelle.
- Les effets de lumière et d'ombre, pour donner de la profondeur et du relief.
- La finesse des textures, qui influent directement sur la crédibilité du rendu.
- Les choix de cadrage et de point de vue, qui orientent la perception.
- L'intégration d'éléments contextuels (personnages, mobilier, végétation) pour ancrer le projet dans son environnement.

[KAIDI, 2021]

PRESENTATION DU CORPUS

Style graphique Emma kaidi

Le mémoire de Emma kaidi est intéressant car elle analyse les styles graphiques des images de synthèses qui ont été réalisés par des grands architectes, et suite à cela elle réalise une liste, selon plusieurs critères technique et esthétique (réalisme, composition de la scène, effets, vue et cadrage...) et elle les nomme.

Collage ou Gondry - Paranormal Activity - Katherine Heigl - Whodunit - Théodore - Mad Max -
Photoréaliste – David

Taxinomie des style graphiques

Style réaliste

Style le plus réaliste et fidèle à la réalité photographique. Il cherche à reproduire les conditions lumineuses, la matérialité et les textures du monde réel avec une extrême précision.



Figure 11. Style photoréaliste par l'agence Foster, Centre de finance à Shanghai en Chine

FIGURE 19 : STYLE REALISTE PAR L'AGENCE FOSTER, CENTRE DE FINANCE A SHANGHAI EN CHINE.

SOURCE : MEMOIRE EMMA KAIDI

Caractéristique selon le tableau de EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping
les effets	ombres portées lumière naturelle effet de réflexion effet de brillance effet de transparence sans saturation
vue et cadrage	vue à hauteur d'œil image en couleur

FIGURE : TABLEAU RECAPITULATIF STYLE PHOTOREALISTE
SOURCE : MEMOIRE EMMA KAIDI

- Degré de réalisme : réaliste
- Composition de la scène : cadrage à hauteur d'œil
- Effets : ombres portées
- Texture : bump mapping
- Lumière : naturelle
- Réflexion : présente
- Brillance : présente
- Transparence : présente
- Image en couleur
- Saturation : sans saturation

Style Paranormal Activity

Style non réaliste qui s'appuie sur une ambiance mystérieuse et surnaturelle. L'image crée une atmosphère étrange à travers une lumière artificielle, des effets visuels accentués (brouillard, halo, contre-jour) et des couleurs froides ou désaturées.



FIGURE 20 : STYLE PARANORMAL ACTIVITY, PROJET DE L'AGENCE DOMINIQUE PERRAULT, POLE INTERMODAL DE GANGNAM, A SEOUL, COREE DU SUD

SOURCE : MEMOIRE DE EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	des billboards : - des personnages - arbres, végétation - mobilier
les effets	transparence collage

FIGURE : TABLEAU RECAPITULATIF STYLE PARANORMAL ACTIVITY
SOURCE : MEMOIRE EMMA KAIDI

Style David

Dérivé du photoréalisme, ce style évoque une perfection artificielle et une ambiance intime. Il est souvent utilisé pour les projets domestiques et se caractérise par des scènes calmes et épurées.



FIGURE 21 :STYLE DAVID PAR L'AGENCE BUTHER, LOW CARBON HOUSE A AVION.
SOURCE: MEMOIRE DE EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping des billboards : -mobilier, véhicule -arbre, végétation
les effets	ombres portées effet de réflexion effet de brillance effet de transparence
vue et cadrage	vue rapprochée vue détaillée image en couleur

FIGURE22 : TABLEAU RECAPITULATIF STYLE DAVID
SOURCE : MEMOIRE EMMA KAIDI

Style MAD MAX

Style expressif et dramatique inspiré des univers cinématographiques post-apocalyptiques. Il mise sur la lumière focalisée et le contraste pour donner une puissance visuelle au projet.



FIGURE 24 : STYLE MAD MAX, PROJET DE ZAHA HADID, MAISON DES ARTS ET DE LA CULTURE ABDULLAH 2 A AMMAN
SOURCE : MEMOIRE DE EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	un contexte (naturel ou urbanisé) des billboards : - arbres, végétation - mobilier
les effets	une lumière focalisée sur le projet
vue et cadrage	vue générale vue en plongée cadrage horizontal

FIGURE 23 : TABLEAU RECAPITULATIF STYLE MADMAX
SOURCE : MEMOIRE EMMA KAIDI

Style Theodore

Style d'équilibre entre réalisme et expressivité. Il traduit la sensibilité du concepteur tout en conservant une rigueur graphique mesurée.



FIGURE 25 : STYLE THEODORE, PROJET DE L'AGENCE CARACTERE SPECIAL, BIBLIOTHEQUE DE LA CITE DES SCIENCES, A PARIS, FRANCE
SOURCE : MEMOIRE DE EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping des billboards : - des personnages - mobilier , véhicule, végétation
les effets	tons chauds lumière naturelle lumière artificielle
vue et cadrage	vue intérieure vue rapprochée

FIGURE 26 : TABLEAU RECAPITULATIF STYLE THEODORE
SOURCE : MEMOIRE EMMA KAIDI

Style Katherine Heigl

Un style photoréaliste poétique, proche du féérique. Il cherche à évoquer la beauté, la douceur et la magie du lieu à travers la lumière et la nature environnante.



FIGURE 27 : STYLE KATHERINE HEIGL, PROJET DE L'AGENCE VEZZONI & ASSOCIES, MUSEE DU BONBON A UZES, IMAGE PRODUITE PAR GOLEM IMAGES

SOURCE : MEMOIRE DE EMMA KAIDI

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	un contexte (naturel ou urbanisé) des billboards : - des personnages - mobilier, véhicule, végétation skybox : coucher de soleil
les effets	tons vifs tons contrastés
vue et cadrage	vue extérieure vue d'ensemble

FIGURE 28 : TABLEAU RECAPITULATIF STYLE KATHERINE HEIGL
SOURCE : MEMOIRE EMMA KAIDI

ETAT DE L'ART DES METHODOLOGIES :

La recherche d'Emma Kaidi (2021) va plus en profondeur sur le style graphique en architecture. Dans son mémoire, elle propose une méthode très directe : elle rassemble un large corpus d'images de concours, définit des critères précis puis construit une grille d'analyse pour classer les rendus selon leur ambiance et leur niveau de réalisme. Elle utilise ensuite Excel pour croiser ces données, ce qui lui permet de faire apparaître rapidement les tendances dominantes.

De leur côté, Beyan & Rossy (2023) ne parlent pas directement de génération d'images : ils cherchent plutôt à cartographier l'évolution de la recherche en IA en s'appuyant sur une analyse bibliométrique des publications. Leur approche permet de comprendre comment progressent les outils et les idées.

Les travaux de « Li et al. (2023) » et « d'Isola et Zhu (2017) » placent l'expérimentation au centre. CycleGAN, Pix2Pix : ils entraînent des modèles sur des dessins d'architectes, puis évaluent si la machine parvient à reproduire le style recherché. Leur protocole repose sur des tests visuels et des analyses directes entre la production de l'IA et l'original.

Marsault & Nguyen (2022) adoptent une autre perspective : ils utilisent les GANs pour stimuler la créativité en phase d'idéation, en jouant avec les réseaux et l'espace latent pour explorer jusqu'où peut aller l'imagination assistée par IA.

Enfin, Rombach et al. (2022) ainsi que Zhang & Liu (2024) se concentrent sur le contrôle des paramètres : variation des prompts, réglage du CFG ou du denoising, et création de multiples formulations pour comprendre l'impact de chaque modification sur le rendu final.

METHODE DE RECHERCHE

La méthode de recherche est l'expérimentation. L'objectif est de déterminer quels sont les paramètres les plus influents sur le style graphique de l'image générée, et cela par degré d'importance. On a donc opté pour une méthode de recherche linéaire, étape par étape :

- En premier, lieux on vas définir notre cadre de recherche, c'est à dire qu'on va choisir d'abord l'outil IA de génération d'image à utiliser et ensuite définir les paramètres de cet outil à analyser.
- Ensuite on procède à l'expérimentation en utilisant cet outil, et ce processus vas suivre une certaine piste. Cette piste consiste à varier chaque paramètre de l'outil un par un en gardant les autres paramètres constants. Et à chaque paramètre on suit la même logique.
- Ensuite les résultats obtenus à travers cette expérimentation seront évalués, l'évaluation se fera en analysant la liste d'image obtenue suivant certains critères qui cible directement l'objectif de cette recherche à savoir le style graphique. Je ferai par la suite une première conclusion pour chaque expérimentation de paramètres. Et une conclusion finale vers la fin.

Maintenant, je vais détailler chaque étape de la recherche pour mieux comprendre le processus de travail :

- 1. Choix de l'outil**
- 2. Processus d'expérimentation**
- 3. Méthode d'Evaluation**

1. Choix de l'outil d'expérimentation

L'outil que j'ai choisis pour cette expérimentation est stable diffusion, la raison de ce choix est que stable diffusion est un outil de génération open source et gratuit pour le grand public. Mais c'est principalement parce que stable diffusion permet à l'utilisateur de contrôler tous les paramètres du modèle de diffusion (denoising, CFG.....). L'utilisateur interagie directement avec les paramètres interne de l'algorithme, on a donc une meilleure compréhension et un meilleur control des résultats obtenu.

Et donc pour bien obtenir des résultats sur stable diffusion il faut comprendre chaque paramètre a quoi il sert contrairement à d'autre outil qui te donne l'impression contrôler le résultat alors qu'il y a des ajustements derrière qui bonifie le résultat pour satisfaire l'utilisateur.

2. Processus d'expérimentation :

L'expérimentation se fait à travers des essais en utilisant stable diffusion, j'ai pris 5 des principaux paramètres de stable diffusion (prompt, image de référence, CFG, denoising, seed) et j'expérimente par la suite sur ces paramètres un par un. Je vais établir par la suite une base des paramètres avec une valeurs constante de chaque paramètre.

À chaque expérimentation je varie un seul paramètre et je garde tout le reste constant pour bien cibler chaque paramètre à lui seul.

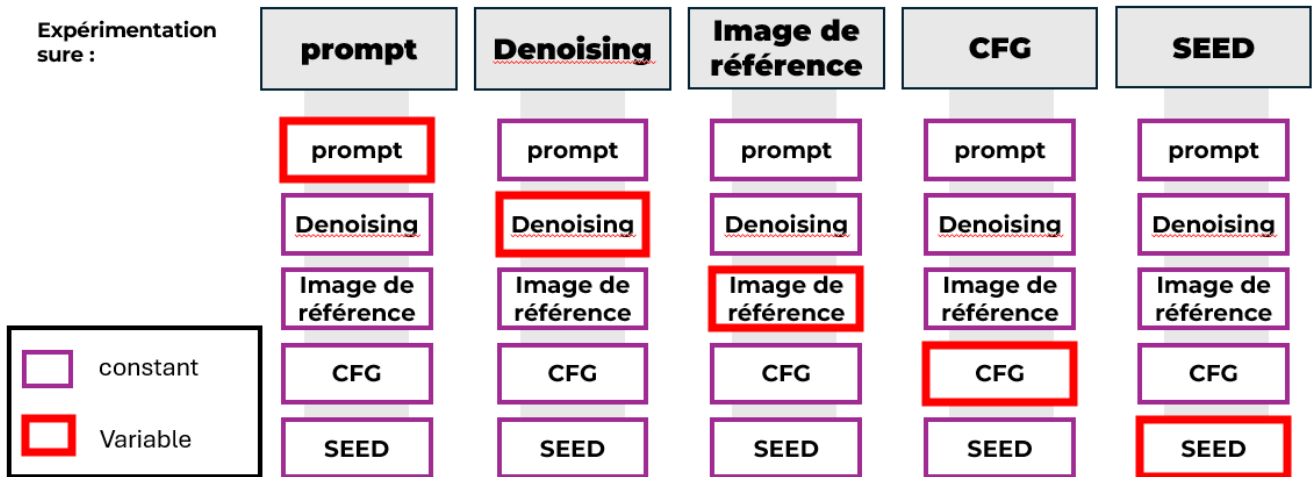


FIGURE 29 : SCHEMA EXPLICATIF DU PROCESSUS DE L'EXPERIMENTATION, DE CHAQUE PARAMETRE
 SOURCE : SCHEMA PERSONNEL

Cette méthode permet de mieux comprendre chaque paramètre indépendamment des autres et donc de mesurer l'impact de chaque paramètre sur le style graphique de l'image.

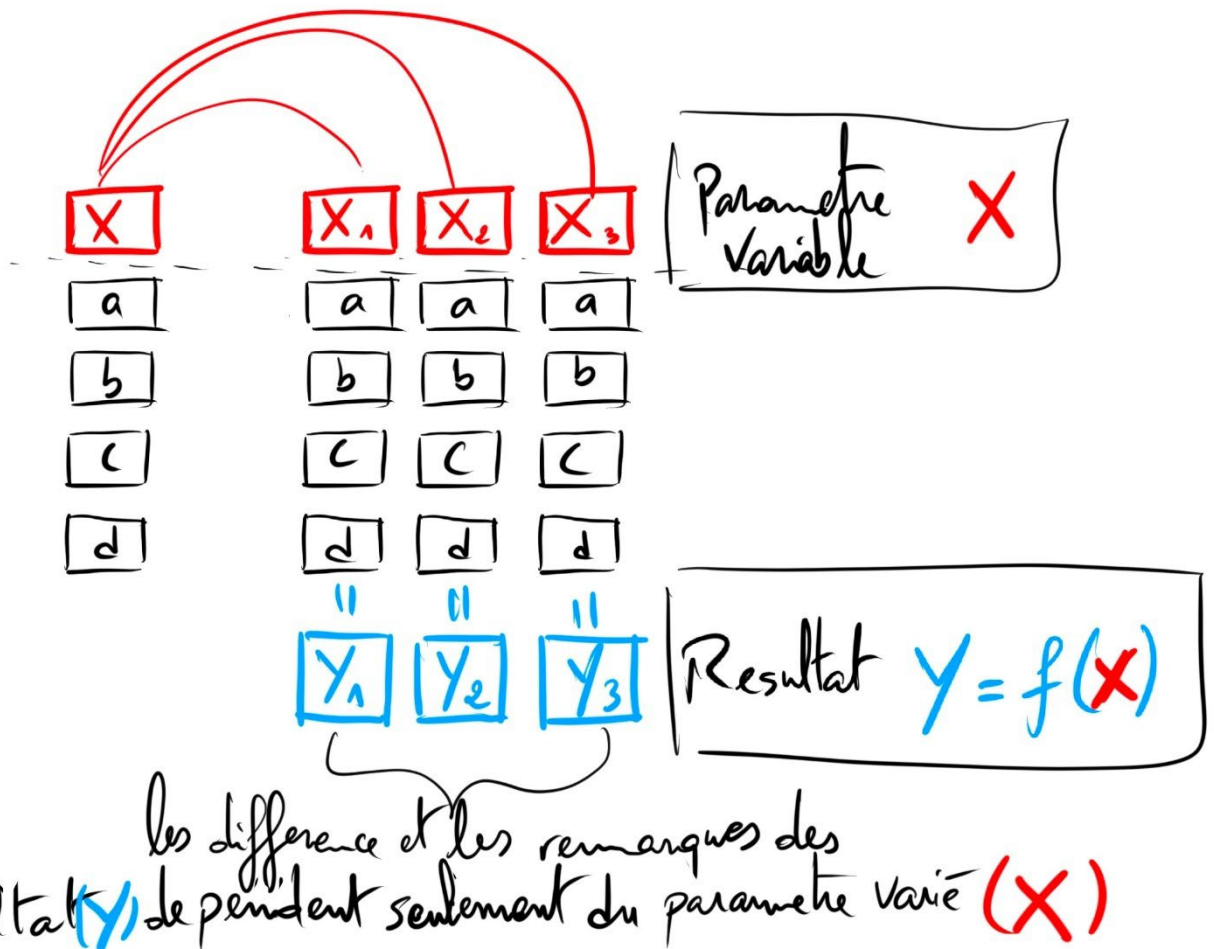


FIGURE 30 : SCHEMA POUR MIEUX COMPRENDRE LE POURQUOI DE CETTE METHODE
 SOURCE : SCHEMA PERSONNEL

A chaque expérimentation on a 4 valeurs constantes, et 1 valeur variable qu'on veut étudier (X). Le résultat obtenu (l'image obtenue (Y)) dépend de chaque paramètres (constant ou variable), et varie en fonction de (X). Donc le résultat obtenu varie en fonction du paramètre qu'on varie, et cela permet de comprendre et de cibler les paramètres un par un. Car si il y a un changement visible sur le résultat, on sait que c'est du à la variante ciblée et pas a autre chose.

Base de données des valeurs constante

L'image de base (controlnet) : on a pris un rendu intérieur généré en utilisant le logiciel de rendu (Lumion)

L'image de référence : aucune



FIGURE 31 : RENDU INTERIEUR D'UNE CHAMBRE GENERE EN UTILISANT LE LOGICIEL LUMION

SOURCE: RENDU PERSONNEL

Prompt standard: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

CFG : 7

Denoising : 50%

SEED : 1 (le bruit ne change pas)

Par la suite je vais expliquer comment on va le varier chaque paramètre :

Expérimentation sur les prompts

L'expérimentation sur les prompts se fera en utilisant des prompts générés par ChatGPT, j'ai choisi ChatGPT car il assure un langage et des formulations optimisées que Stable Diffusion interprète efficacement. Je lui donnerai des directives bien précises.

Elle se fera en deux parties, la première en variant l'ordre d'écriture des mots du prompt, la deuxième en variant le prompt par rapport à des styles graphiques.

Pour cela je prends comme référence le style graphique que Emma Kaidi a nommé « style réaliste » et je formule un prompt selon les caractéristiques de ce prompt (toujours en se basant sur les informations tirées du mémoire),

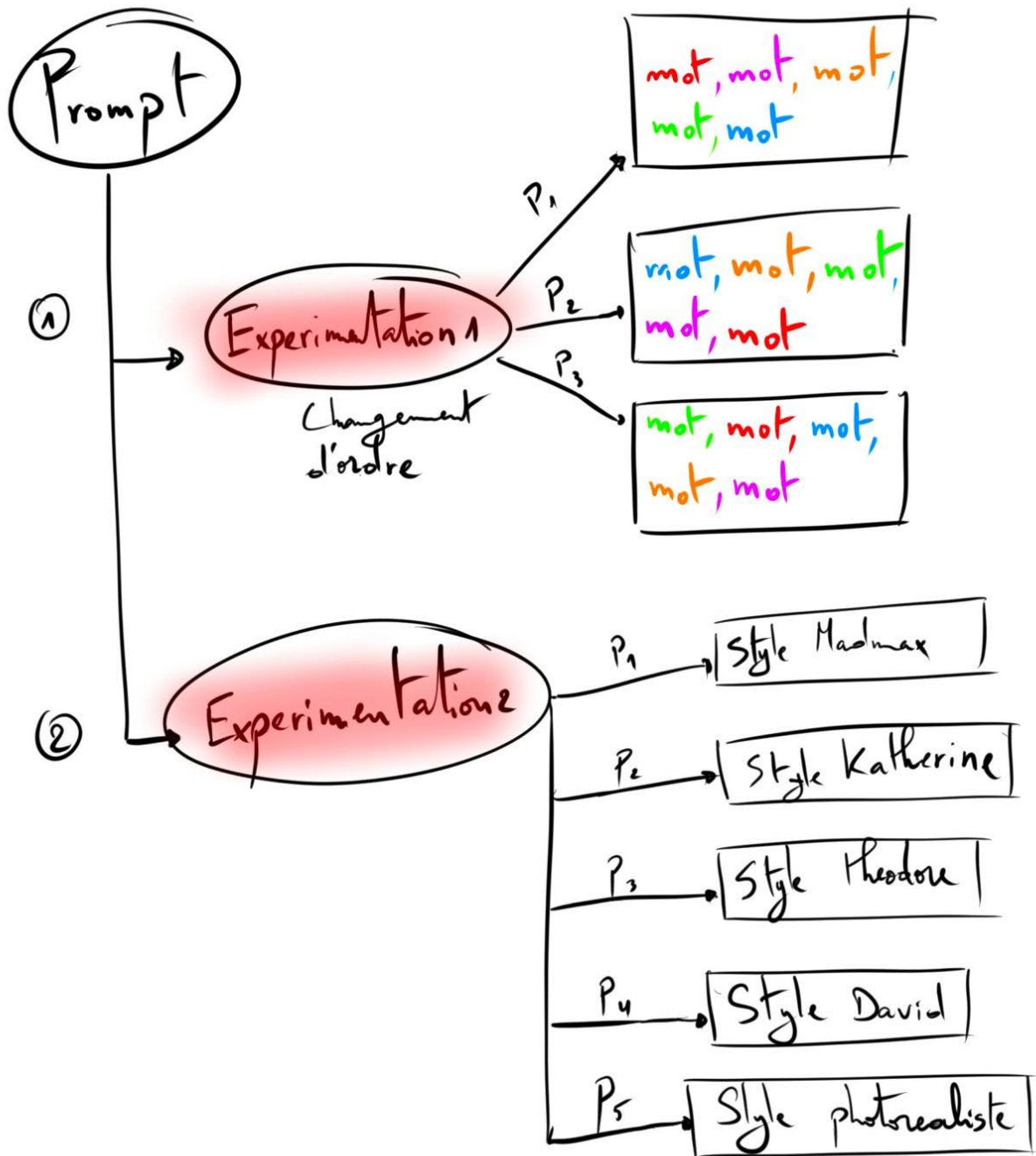


FIGURE 32 : SCHEMA EXPLIQUANT LA METHODE DE RECHERCHE SUR L'EXPERIMENTATION DES PROMPTS
 SOURCE : SCHEMA PERSONNEL

Les informations sont tirées du tableau de emma kaidi qui contient les paramètres graphiques qu'elle a identifié, et elle résume ce tableau ci-dessous.

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping
les effets	ombres portées lumière naturelle effet de réflexion effet de brillance effet de transparence sans saturation
vue et cadrage	vue à hauteur d'oeil image en couleur

FIGURE 33 : TABLEAU RECAPITULATIF DES CRITERE DU STYLE PHOTOREALISTE
SOURCE : (EMMA KAIDI)

Voici le prompt que j'ai écrit :

« *Un rendu réaliste d'un espace intérieur, vue à hauteur d'œil, lumière naturelle, ombres portées, texture bump mapping, effet de réflexion, effet de brillance, effet de transparence, image en couleur, sans saturation* »

Voici la traduction en anglais pour l'utiliser sur stable diffusion :

"a realistic rendering of an interior space, eye-level view, natural light, cast shadows, bump mapping texture, reflection effect, gloss effect, transparency effect, color image, without saturation"

Ensuite, j'ai donné le prompt à chatgpt pour qu'il me genere un prompt plus efficace pour stable diffusion, voici ce que j'ai écrit a chatgpt :

Voici le prompt que j'ai écrit :

« *"Realistic rendering of a building, eye-level view, natural light, cast shadows, bump mapping texture, reflection effect, gloss effect, transparency effect, color image, without saturation"*

Donne-moi le prompt le plus adapté pour stable diffusion »

Voici le résultat que chatgpt m'a donné :

"Ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization"

Avec cette méthode j'obtiens mon premier prompt, je ferais de même pour les autres styles d'emma kaidi par la suite.

Expérimentation 1 (Ordre des mots)

En premier lieux j'ai essayé de voir si l'ordre des mots influence directement le résultat obtenu

Pour cette partie on change donc que l'ordre d'écriture pour voir si cela exerce une influence sur les résultats sur les 3 variations

Variation 1: prompt de base:

"Ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces,

subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization”

Pour les autres variations on prend les passages que j’ai mis en couleur (chaque passage évoque un détail ou une information qui sera reçu par stable diffusion) Et on change l’ordre des passages. Comme montré sur le schéma ci-dessous.

“Ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization”

viewed at human eye level, ultra realistic architectural rendering of a modern building, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization

FIGURE 34 : SCHEMA MONTRANT LES DEUX PROMPTS EN CHANGEANT L'ORDRE D'ECRITURE

SOURCE: SCHÉMA PERSONNEL

Variation 2:

“Viewed at human eye level, ultra realistic architectural rendering of a modern building, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization”

Variation 3:

“Detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, full color image, low saturation, physically accurate lighting, high-quality 3D visualization”

Experimentation2 (variations de prompts selon un style graphique bien défini)

Je prends 5 styles graphique les plus expressif de Emma kaidi et Je génère les prompt de chaque style de la même manière que j’ai fait avec le premier (à partir du tableau récapitulatif du style graphique) voici donc

chaque style graphique avec son tableau récapitulatif et le prompt que j'ai généré à partir de ce tableau ainsi que le prompt généré ensuite avec chatgpt :

Style photorealist

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping
les effets	ombres portées lumière naturelle effet de réflexion effet de brillance effet de transparence sans saturation
vue et cadrage	vue à hauteur d'oeil image en couleur

FIGURE 35 : TABLEAU RECAPITULATIF STYLE PHOTOREALISTE
SOURCE : MEMOIRE EMMA KAIDI

Prompt personnel : Un rendu réaliste d'un espace intérieur, vue à hauteur d'œil, lumière naturelle, ombres portées, texture bump mapping, effet de réflexion, effet de brillance, effet de transparence, image en couleur, sans saturation.

Prompt chatgpt:

"Ultra realistic architectural rendering of a modern space, interior view at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization."

Style paranormal activity:

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input checked="" type="checkbox"/> non réaliste
composition de la scène	des billboards : - des personnages - arbres, végétation - mobilier
les effets	transparence collage

FIGURE 36 : TABLEAU RECAPITULATIF STYLE PARANORMAL ACTIVITY
SOURCE : MEMOIRE EMMA KAIDI

Prompt personnel : un rendu non réaliste d'un bâtiment, lumière artificielle, contrastes forts, halo lumineux, couleurs froides, atmosphère étrange, ombres déformées, texture simple, vue décentrée

Prompt chagpt:

“surreal architectural rendering of a building, artificial lighting with strong contrasts and glowing halos, cold color palette, eerie atmosphere, distorted shadows, simple flat textures, off-centered composition, cinematic mysterious mood, non-realistic style.”

Style mad max :

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	un contexte (naturel ou urbanisé) des billboards : - arbres, végétation - mobilier
les effets	une lumière focalisée sur le projet
vue et cadrage	vue générale vue en plongée cadrage horizontal

FIGURE 37 : TABLEAU RECAPITULATIF STYLE MADMAX
SOURCE : MEMOIRE EMMA KAIDI

Prompt personnel : un rendu semi-réaliste d'un bâtiment, vue aérienne, lumière focalisée sur le projet, contexte sombre, ciel nuageux, ombres fortes, couleurs désaturées, ambiance dramatique

Prompt chagpt:

“semi-realistic aerial architectural rendering, dark post-apocalyptic atmosphere, focused spotlight on the building, cloudy sky, deep shadows, desaturated tones, dramatic mood, cinematic lighting, high-contrast composition”

Style Katherine heighl :

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	un contexte (naturel ou urbanisé) des billboards : - des personnages - mobilier, véhicule, végétation skybox : coucher de soleil
les effets	tons vifs tons contrastés
vue et cadrage	vue extérieure vue d'ensemble

FIGURE 38 : TABLEAU RECAPITULATIF STYLE KATHERINE
SOURCE : MEMOIRE EMMA KAIDI

Prompt personnel : Un rendu réaliste d'un projet architectural, lumière douce et diffuse, végétation dense et expressive, ambiance bucolique, teintes chaudes et naturelles, composition harmonieuse, présence d'eau et de matière minérale

Prompt chagpt:

"Realistic architectural rendering, soft diffuse light, expressive lush vegetation, warm natural tones, bucolic and poetic atmosphere, harmonious and balanced composition, presence of water and stone materials, soft lighting and realistic environment."

Style Theodore:

Les caractéristiques de l'image	Les critères
dégrés de réalisme	<input checked="" type="checkbox"/> réaliste <input type="checkbox"/> non réaliste
composition de la scène	texture : bump mapping des billboards : - des personnages - mobilier , véhicule, végétation
les effets	tons chauds lumière naturelle lumière artificielle
vue et cadrage	vue intérieure vue rapprochée

FIGURE 39 : TABLEAU RECAPITULATIF STYLE THEODORE
SOURCE : MEMOIRE EMMA KAIDI

Prompt personnel : un rendu réaliste et expressif d'un bâtiment, éclairage doux, détails précis mais modérés, texture réaliste, ambiance sensible, composition équilibrée, couleurs naturelles.

Prompt chagpt:

“Realistic yet expressive architectural rendering, soft lighting, moderate level of detail, realistic textures, emotionally sensitive atmosphere, balanced composition, natural color palette, artistic yet technically precise visualization.”

Expérimentation sur le CFG :

Le traitement du CFG est assez particulier car il est directement lié aux prompt (CFG : degré de fidélité du modèle à suivre le prompt) et donc pour ce paramètre on va varier le CFG, en allant de 1 à 30

- CFG = 1
- CFG = 5
- CFG = 10
- CFG = 20
- CFG = 30

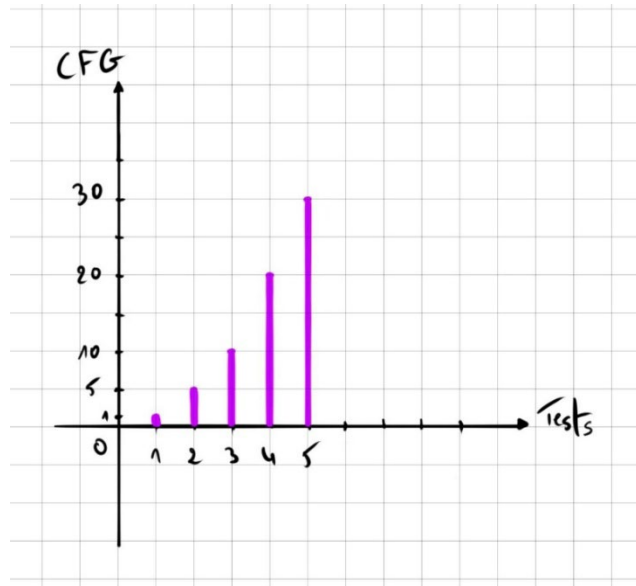


FIGURE 40 : DIAGRAMME DE L'EXPERIMENTATION SUR LE CFG
SOURCE : SCHEMA PERSONNEL

le diagramme sert juste à visualiser les chiffres que je vais mettre comme donnée du cfg et de comparer par la suite quand on aura les résultats.

Expérimentation Denoising strength :

Variation du denoising de 0% à 90%, donc

- 10%
- 20%
- 30%
- 40%
- 50%
- 60%
- 70%
- 80%
- 90%

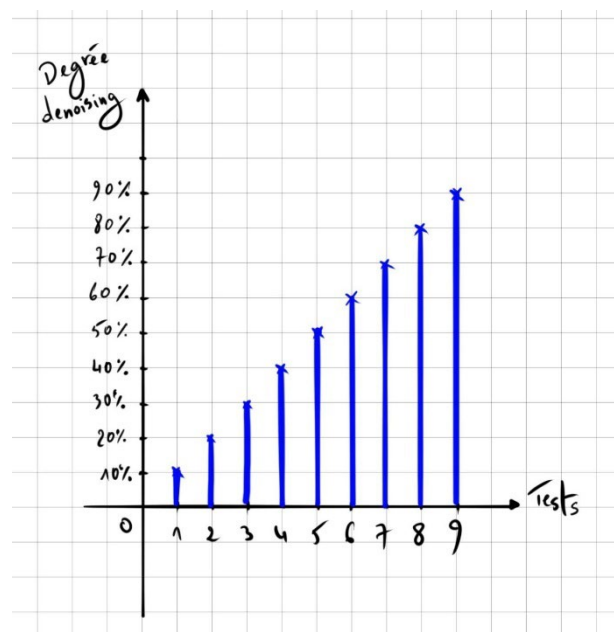


FIGURE 41 : DIAGRAMME DE L'EXPERIMENTATION SUR LE DENOISING
SOURCE : SCHEMA PERSONNEL

Le diagramme sert juste à visualiser les chiffres que je vais mettre comme donnée du denoising et de comparer par la suite quand on aura les résultats.

Expérimentation Image de référence :

Pour cette partie j'ai choisi plusieurs images de référence qui évoque un certain style graphique particulier, je me suis référencé aux mémoire d'EMMA KAIDI, car dans son mémoire elle a bien analysé les styles graphiques des images de synthèse, et donc choisir des styles graphiques qui ont fait l'objet d'une étude est plus intéressant pour ma recherche car on a des styles graphiques qui sont déjà étudié et analysé.

On va aussi expérimenter avec des images de référence qui n'ont pas de contenu architectural ou autres (image texturé), pour voir si cela a un impact aussi ou pas.

Expérimentation SEED :

SEED= -1 (seed aléatoire) cela veut dire qu'à chaque fois le bruit sera différent du précédant.

On verra si le seed change et est ce qu'il y'aura un impact sur le graphisme de l'image finale.

Méthode d'évaluation des résultats

La méthode consiste à analyser les résultats, expérience par expérience et faire des conclusions pour chaque expérimentation effectuée. Et je vais faire vers la fin une conclusion final.

RESULTATS DE L'EXPERIMENTATION :

L'objectif pour le moment c'est de constituer une base de données d'images générées avec stable diffusion. Il ne reste plus qu'à appliquer en variant un par un les paramètres.

VARIATION DE PROMPT

1. Variation de l'ordre d'écriture :

Hypothèse de base 2 : est-ce que l'ordre des mots exerce une importance sur le style graphique.

Prompt1: Ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization

Denoising 40%

CFG 7

SEED 1

ControlNet/image de référence = identique



Prompt1: viewed at human eye level, ultra realistic architectural rendering of a modern building, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization

Denoising 40%

CFG 7

SEED 1

ControlNet/image de référence = identique



Prompt1: detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, ultra realistic architectural rendering of a modern building, viewed at human eye level, under natural daylight, soft cast shadows, full color image, low saturation, physically accurate lighting, high-quality 3D visualization

Denoising 40%

CFG 7

SEED 1

ControlNet/image de référence = identique





Analyse prompts 1

Les résultats montrent que la modification de l'ordre des mots dans le prompt n'entraîne aucun changement perceptible sur le rendu final. Les trois variations ont produit des images quasiment identiques, tant au niveau de la composition, des textures, de la lumière que des détails. Cela confirme que Stable Diffusion ne tient pas compte de la structure des phrases, mais se base plutôt sur les mots-clés.

Conclusion prompts 1

je conclut donc que l'ordre d'écriture n'a aucun impact sur le contenu de l'image finale, encore moins sur le style graphique, ce qui répond à l'hypothèse 2.

2. Variation de prompt à partir d'un style graphique bien défini

Prompt1: (Style photorealiste)

"Ultra realistic architectural rendering of a modern space, interior view at human eye level, under natural daylight, soft cast shadows, detailed bump-mapped textures, realistic reflections, smooth glossy surfaces, subtle transparency, full color image, low saturation, physically accurate lighting, high-quality 3D visualization."

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique



Prompt1: (Style paranormal)

"surreal architectural rendering of a building, artificial lighting with strong contrasts and glowing halos, cold color palette, eerie atmosphere, distorted shadows, simple flat textures, off-centered composition, cinematic mysterious mood, non-realistic style."

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique





Prompt1: (Style madmax)

"semi-realistic aerial architectural rendering, dark post-apocalyptic atmosphere, focused spotlight on the building, cloudy sky, deep shadows, desaturated tones, dramatic mood, cinematic lighting, high-contrast composition"

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique



Prompt1: (Style Katherine heighl)

"Realistic architectural rendering, soft diffuse light, expressive lush vegetation, warm natural tones, bucolic and poetic atmosphere, harmonious and balanced composition, presence of water and stone materials, soft lighting and realistic environment."

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique



Prompt1: (Style Théodore)

"Realistic yet expressive architectural rendering, soft lighting, moderate level of detail, realistic textures, emotionally sensitive atmosphere, balanced composition, natural color palette, artistic yet technically precise visualization."

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique





Analyse prompts 2

En comparant les rendus produits, les différences que j'ai observées sont présentes mais relativement faibles.

Chaque image conserve une structure et une ambiance générale similaires, avec seulement de légères variations dans la lumière, les teintes et la densité des détails. Les changements dans le style graphique sont donc perceptibles mais peu marqués. En résumé, le style graphique global reste le même dans tous les 5 photo, malgré les variations de formulation et d'intention stylistique, stable diffusion suit donc le style graphique de l'image de base en priorité et ajoute des retouche a partir du prompt (il n'exagère pas dans l'interprétation du prompt (en tout cas sur les mots clé qui font référence aux style graphique))

Conclusion prompts 2


Je conclus donc que le prompt à une très légère influence sur le style graphique de l'image finale.

Variation de Denoising:

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	0%
CFG	7
SEED	1

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	10%
CFG	7
SEED	1


ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	20%
CFG	7
SEED	1

ControlNet/image de référence = identique





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising

30%

CFG

7

SEED

1

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising

40%

CFG

7

SEED

1

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising

50%

CFG

7

SEED

1

ControlNet/image de référence = identique





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	60%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	70%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	80%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------





Analyse denoising :

Les images obtenues avec différentes valeurs de *denoising* présentent des changements très progressifs mais significatifs.

- Pour des valeurs entre (0.1 à 0.3), les rendus sont très proches de l'image de base les couleurs, la lumière et la composition sont quasiment identiques.
- À partir de 0.5, je remarque un changement plus visible du style graphique, changement de lumière, texture, couleurs. Néanmoins j'ai observé aussi des changements dans les détails et le contenu par rapport à l'image de base
- Entre 0.7 et 0.9, je remarque un changement assez conséquent dans le style graphique ms aussi une total transformation dans le contenu et les détails par rapport à l'image de base

Conclusion denoising :

Cette expérience montre donc que le *denoising strength* a une influence sur le style graphique, néanmoins le résultat perd des détails et des éléments de l'image de base jusqu'à avoir une image totalement différente.

Plus on a un denoising plus grand plus il y a un changement sur le style graphique moins l'image est réaliste et vice versa

VARIATION DE l'image de référence :

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising 40%

CFG 7

SEED 1

ControlNet



image de référence



Theodore

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising 40%

CFG 7

SEED 1

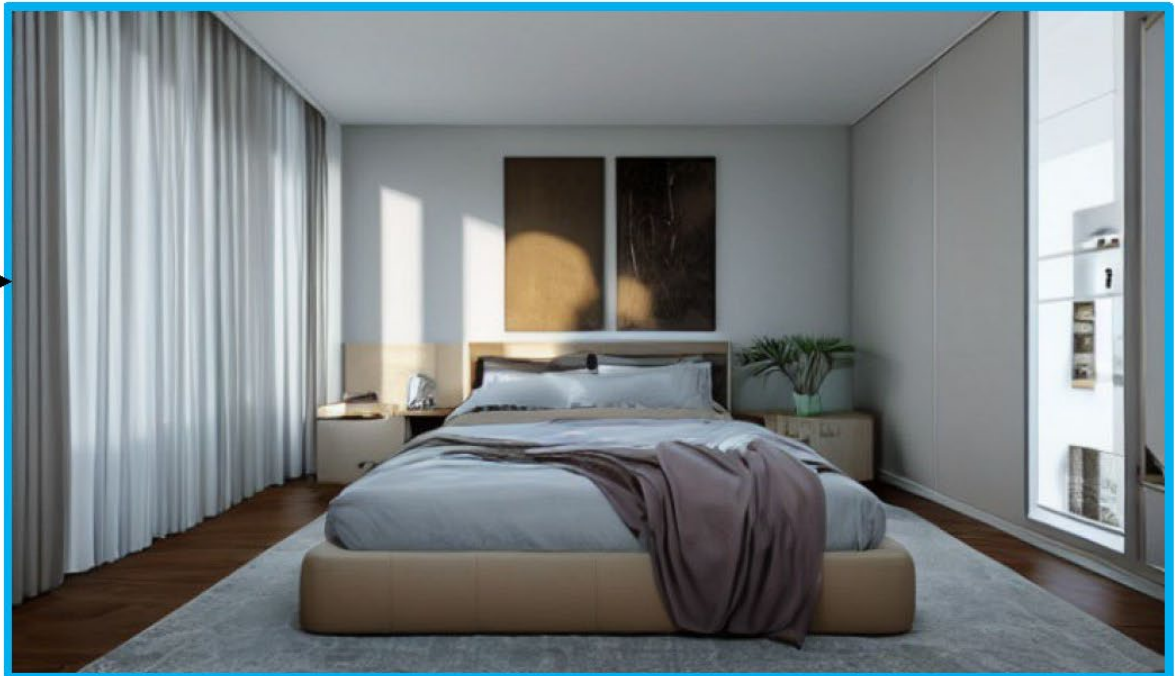
ControlNet



image de référence



David



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	40%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------



image de référence



Katherine heigl

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	40%
------------------	------------

CFG	7
------------	----------

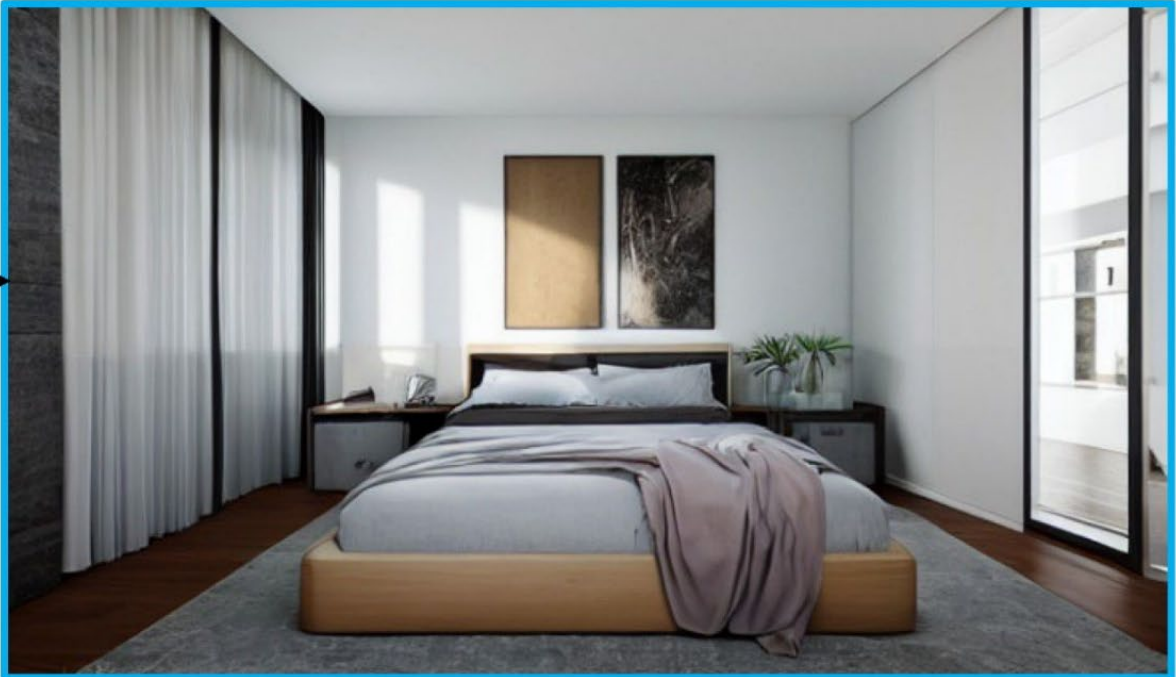
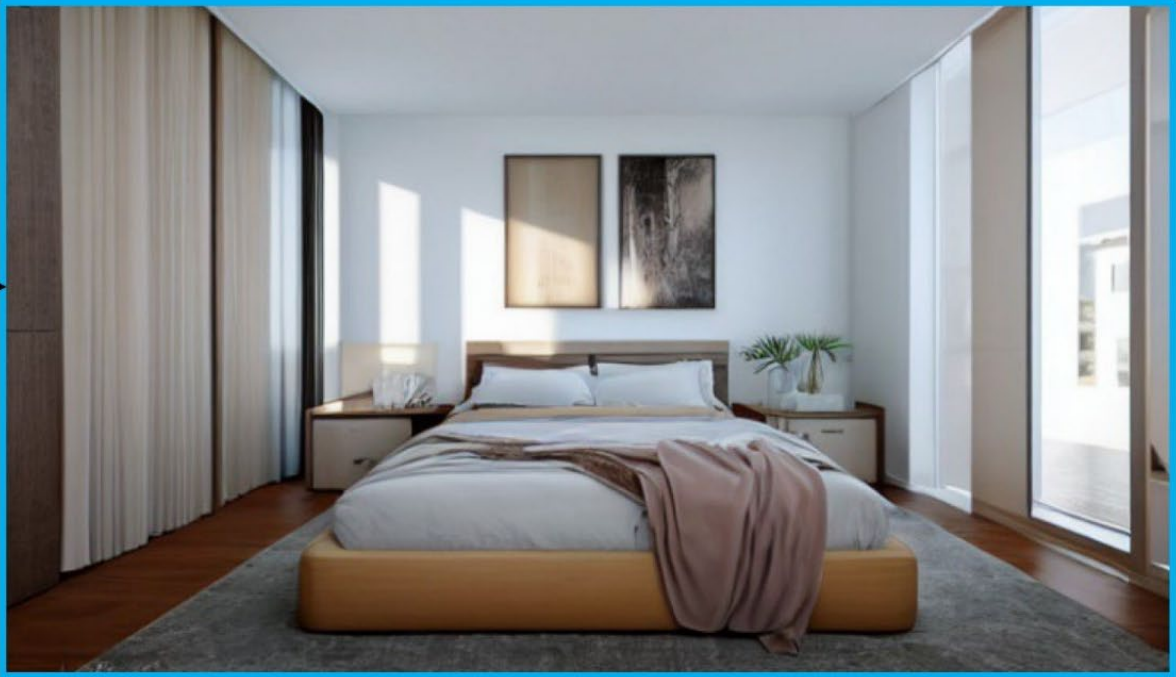
SEED	1
-------------	----------



image de référence



Paranormal activity



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	40%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	40%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------



image de référence



Photo pinterest

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	40%
------------------	------------

CFG	7
------------	----------

SEED	1
-------------	----------



image de référence



Texture pinterest



Analyse image de référence :

Les résultats obtenus montrent un changement assez conséquent sur le style graphique des images générées, la structure globale de l'image reste la même (il n'y a pas de différence flagrante), que de légers changements sur les détails et éléments de l'espace (éléments décoratifs, ajout ou suppression de mobilier par exemple).

Les images issues de références architecturales du travail de Emma Kaiji (les 5 premiers) conservent la structure spatiale et la logique lumineuse du modèle de départ (elles sont plus fidèles à l'image de base)

Les images utilisant des références non architecturales (la dernière) produisent des résultats un peu moins fidèles à l'image de base, elles sont légèrement abstraites par rapport aux 5 premiers.

La différence la plus flagrante est aux niveaux des couleurs, les images architecturales contiennent une palette de couleurs plus diversifiée, ce qui donne à l'image générée un aspect plus vivant, contrairement à l'image abstraite

Conclusion image de référence :

Je peux donc conclure que l'image de référence influence considérablement le style graphique de l'image générée.

Variation de CFG :

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
CFG	1
SEED	1


ControlNet/Image de référence = Identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
CFG	5
SEED	1

ControlNet/Image de référence = Identique

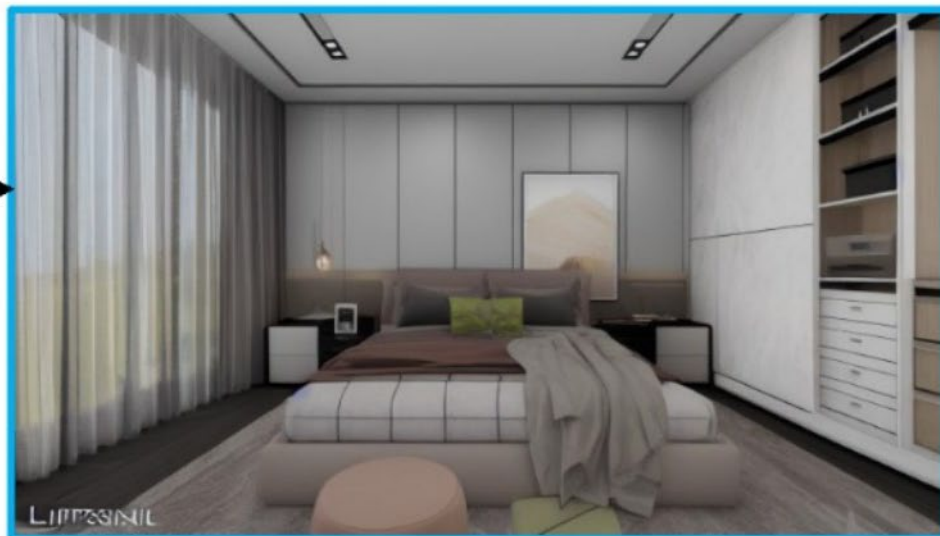


Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
CFG	10
SEED	1

ControlNet/Image de référence = Identique





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising

90%

CFG

20

SEED

1

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising

90%

CFG

30

SEED

1

ControlNet/image de référence = identique





Analyse CFG :

Les résultats obtenus montrent de léger changement sur le style graphique sur les l'images générer.

- Pour les valeurs basses (CFG = 1 à 5), il n'y a quasiment aucun changement sur le contenu de l'image ni sur le style graphique.
- À partir de (CFG à 10), le style graphique commence a changer légèrement par rapport a l'image de base, je ne remarque pas non plus une grande différence sur les éléments de détails de l'image. Pour la valeur la plus hautes (CFG = 30), l'image obtenue montre un changement encore plus grand sur le style graphique mais pas un très grand changement non plus, avec un accentuent sur les textures, les couleurs et la lumière.

Il n'y a pas de changements flagrants sur les détails de l'image. Néanmoins l'image obtenu parait moins naturelle (couleur saturé, trait durcit, moins fidèle à l'image de base) avec un degré de CFG de ce niveau.

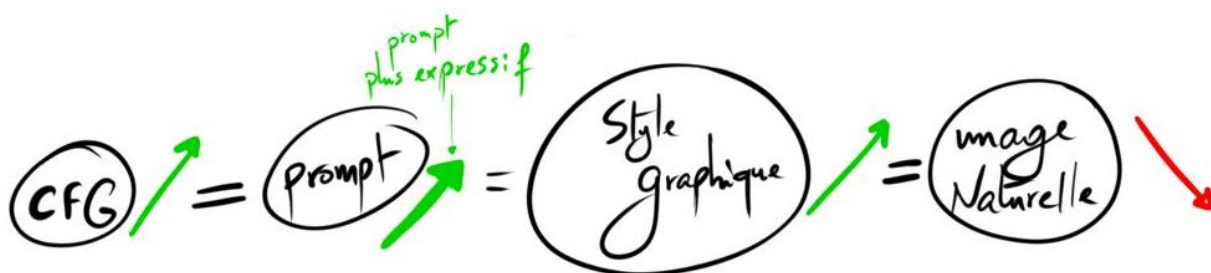


FIGURE 42 : SCHEMA RESUMANT L'ANALYSE DU SEED

SOURCE : schéma personnel

Conclusion CFG :


Je peux conclure donc que le CFG influence surtout le degré de fidélité au prompt qui lui influence légèrement sur le style graphique, néanmoins un degré de CFG trop élevée dénature un peu l'image généré.

Variation de seed :

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
CFG	20
SEED	70


ControlNet/image de référence = identique

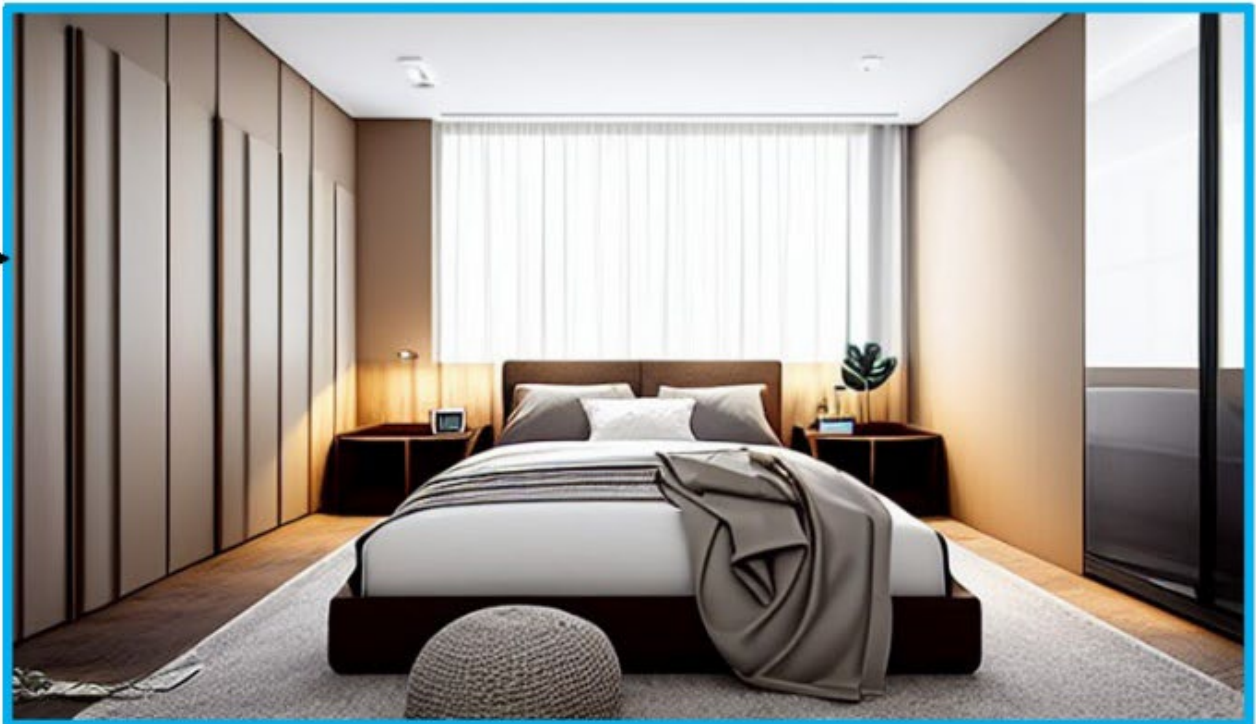
A photograph of a modern bedroom. It features a bed with a white and brown duvet, a large window with sheer curtains, and minimalist furniture. The room is brightly lit with natural light. A small 'L. Jumien' logo is visible in the bottom left corner of the image.

Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
CFG	20
SEED	70

ControlNet/image de référence = identique

A photograph of a modern bedroom, identical to the one in the first block. It features a bed with a white and brown duvet, a large window with sheer curtains, and minimalist furniture. The room is brightly lit with natural light. A small 'L. Jumien' logo is visible in the bottom left corner of the image.



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
------------------	------------

CFG	20
------------	-----------

SEED	1234
-------------	-------------

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
------------------	------------

CFG	20
------------	-----------

SEED	1234
-------------	-------------

ControlNet/image de référence = identique





Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

Denoising	90%
------------------	------------

CFG	20
------------	-----------

SEED	2
-------------	----------

ControlNet/image de référence = identique



Prompt1: "A modern bedroom with large windows, wooden flooring, minimalist furniture, neutral color tones, soft natural light, indoor plants by the window, and a neatly made bed with linen sheets. Cozy and serene atmosphere."

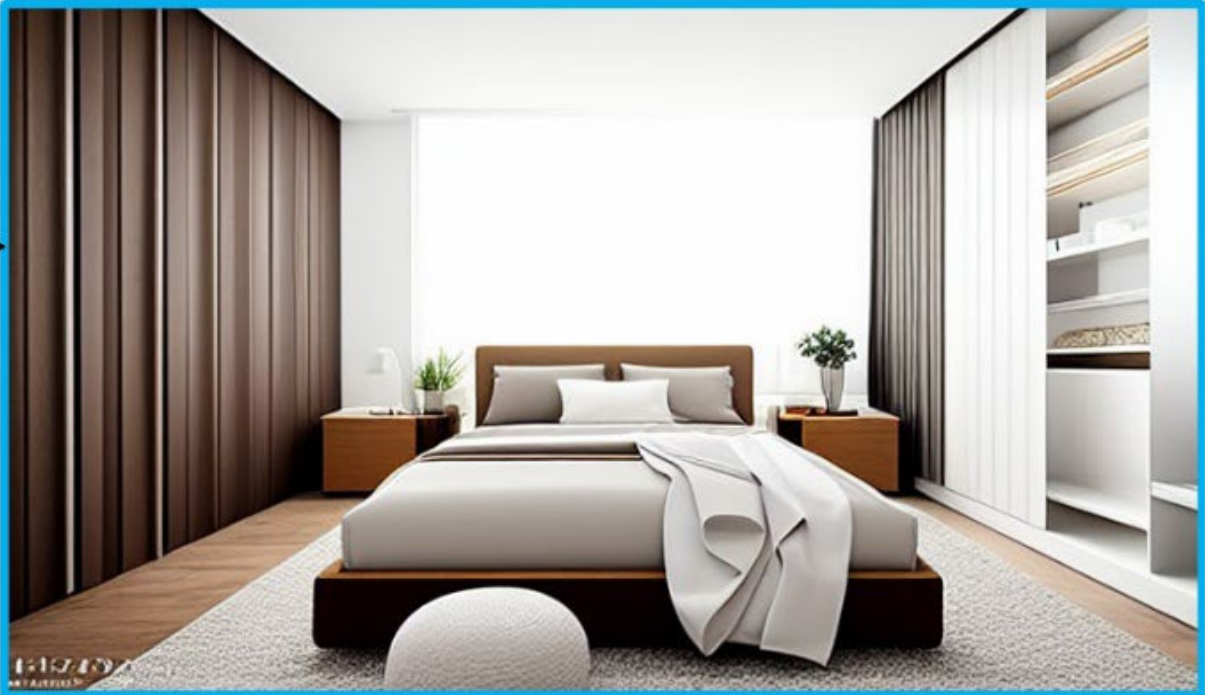
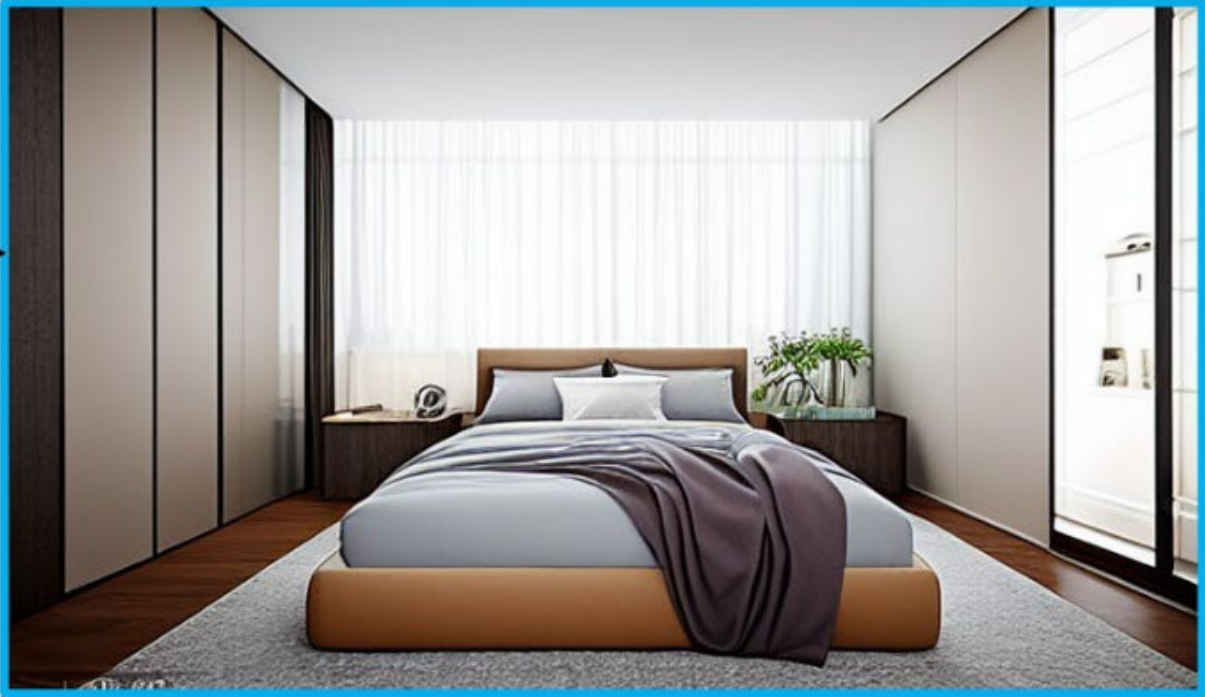
Denoising	90%
------------------	------------

CFG	20
------------	-----------

SEED	3
-------------	----------

ControlNet/image de référence = identique





Analyse SEED :

Les images générées avec les même SEED montrent des résultats identiques a chaque fois. Que ce soit sur le style graphique ou sur le contenu des images. Et ce quel que soit le nombre du SEED de 0 a l'infinie.

En revanche les images générées avec différent SEED montre de légers changements sur certains détails du contenu et mais très léger sur le style graphique global de l'image, c'est juste dans la répartition de la lumière, des couleurs....

Le numéro de SEED n'a pas d'influence direct sur le style graphique ou le contenu, par exemple un SEED très grand (1234) ne veut pas dire plus d'élément graphique ou style graphique plus esthétique, le SEED donne juste un bruit qui limite la composition de l'image générée (chaque numéro de 0 a l'infinis a son propre bruit), si le SEED reste le même l'image reste exactement la même. Si le bruit (SEED) change, certains détails de l'image change comme les éléments, le décor, les couleurs changent légèrement,

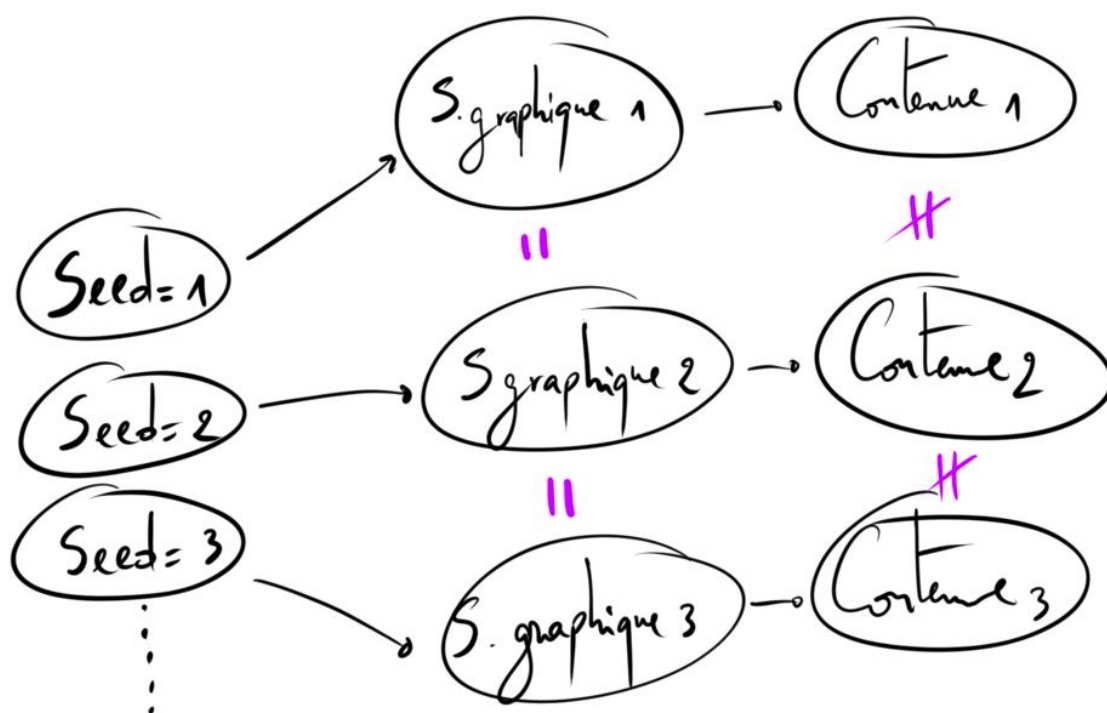


FIGURE 43 : SCHEMA RESUMANT L'ANALYSE DU SEED

SOURCE : schéma personnel

Conclusion SEED :

Donc je ne peux pas dire que le SEED a une influence directe sur le style graphique car le changement qu'apporte un SEED différent sont des éléments de contenu et qui sont aléatoire (que l'utilisateur ne contrôle même pas).

Confrontation avec les hypothèses

Hypothèse 1 : La précision lexicale dans la formulation des prompts influences directement le style graphique de l'image générée.

Hypothèse 2 : est-ce que l'ordre des mots exerce une importance sur le style graphique.

Avec l'expérimentation sur les prompts j'ai pu répondre aux deux hypothèses, en premier lieu, la précision lexicale dans la génération du prompt a une légère influence sur le style graphique. En deuxième lieu, l'ordre d'écriture des mots dans les prompts n'a aucune influence sur le style graphique.

SYNTHESE GLOBALE

Dans cette synthèse je vais classer les paramètres du plus important au moins important par rapport au style graphique, en se basant sur les résultats d'expérimentation que j'ai fait voici la synthèse finale que j'ai tiré.

L'Image de Référence

C'est le paramètre le plus important et à prendre en considération en premier.

L'expérimentation montre que Stable Diffusion suit en priorité la structure de l'image de ControlNet, et la logique lumineuse de l'image de référence. C'est cette dernière qui définit la palette de couleurs, l'ambiance de l'image (Style graphique). En Image-to-Image, l'image de référence verrouille et définit le style, l'IA quand elle agit comme une retouche sur l'image et pas comme créateur indépendant.

Le Denoising Strength

Le denoising a un impact assez significatif et progressif.

Néanmoins, il y a un rapport inverse entre le style graphique et la fidélité du contenu de l'image de base, (plus le style graphique change, moins l'image est fidèle). Si on veut accentuer le style graphique qu'on veut avoir, on perd en termes de fidélité du contenu de l'image de base, jusqu'à obtenir une image totalement irréaliste. Donc il y a un équilibre à gérer entre style et réalité.

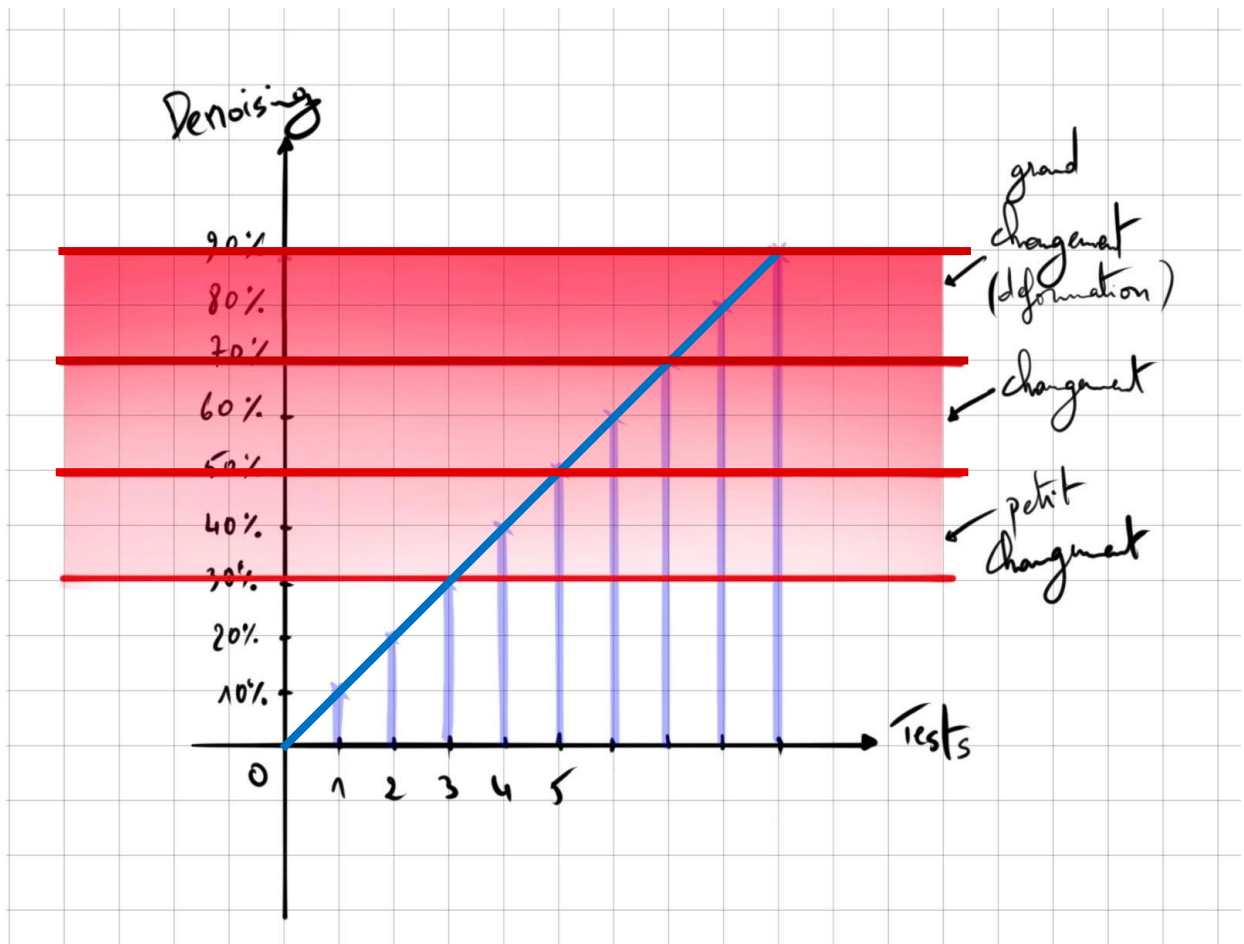


FIGURE 44 : GRAPHE MONTRANT LES DIFFERENTS NIVEAUX CHANGEMENT APORTE EN FONCTION DE DENOISING

SOURCE : graphique personnel

Le CFG Scale

Le cfg agit comme un amplificateur, son impact est Modéré voir faible.

Il influence surtout le degré de fidélité au prompt, il ne crée pas le style graphique en lui seul. Et Tout comme le denoising, son impact sur style graphique n'est visible qu'à des valeurs assez grandes. Néanmoins, et contrairement aux denoising, le résultat ne devient pas irréel, et il n'y a pas de grande déformation visible dans l'image (la structure et le contenu reste le même), les différences sont surtout sur la saturation des couleurs, traits durcis ce qui rend l'image moins naturelle.

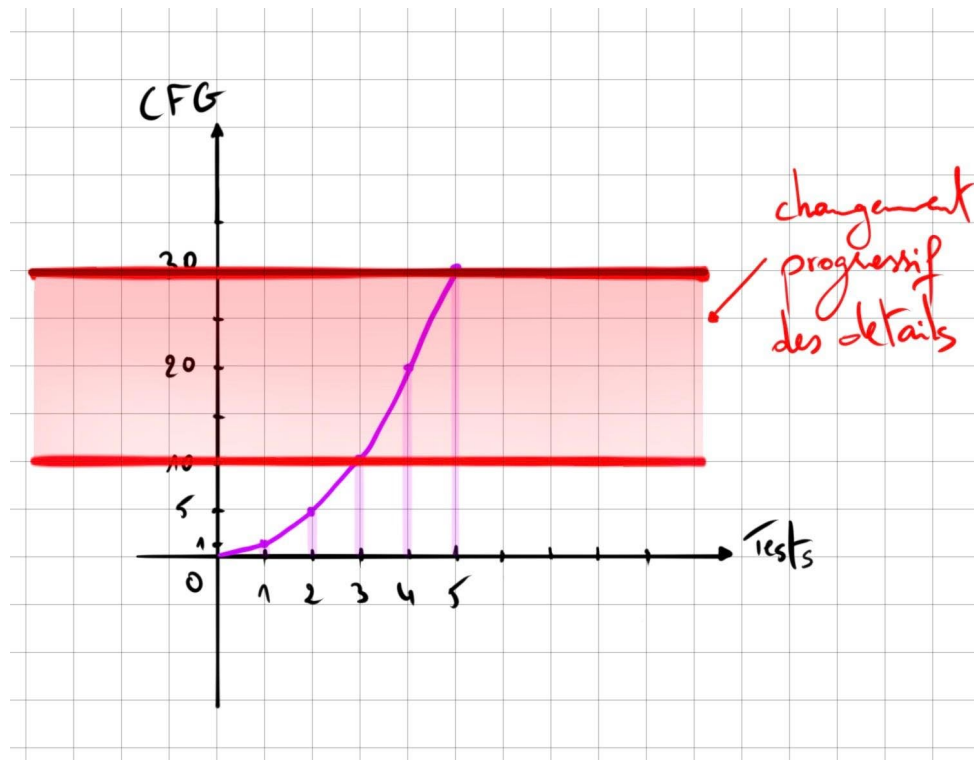


FIGURE 45 : SCHEMA RESUMANT L'ANALYSE DU SEED

SOURCE : schéma personnel

Le Prompt

Le prompt a un impact très léger.

L'ordre des mots n'a aucun impact et Les variations de formulation n'apportent de grand changement graphique non-plus

Relation Prompt / Image de référence : L'IA privilégie le style graphique de l'image de référence par rapport aux instructions du prompt. Le prompt n'est qu'un complément d'information.

Le SEED

Le SEED a un impact très faible et aléatoire.

Il n'influence pas le style mais la répartition du bruit qui lui influence les détails et le décor. Et ce changement est aléatoire et (pas de contrôle de l'utilisateur). C'est un outil qui stabilise le résultat s'il ne change pas et change aléatoirement et très faiblement le style en changeant, ce n'est pas un outil de création de style

Paramètre	Dégré d'importance pour le style graphique
Prompt	
Denoising	
Image de reference	
CFG	
SEED	

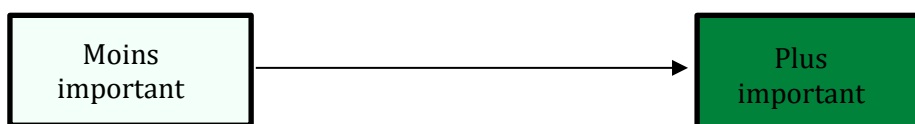


TABLEAU RECAPITULATIF (DEGRE D'IMPORTANCE PAR CODE COULEUR)

Limites et perspectives

1. Les Limites du travail

Malgré les résultats produits de cette expérimentation, ce travail présente certaines limites :

La partie subjectivité : l'évaluation d'un "style graphique" comporte une petite part de subjectivité. Ce qui est perçu comme "réaliste" ou "artistique" peut varier selon l'observateur, et aussi il n'y a pas de métrique précise pour évaluer les résultats et donner des résultats chiffrés, par exemple de mettre à une échelle de 1 à 10 par degré d'importance,

Le matériel informatique (Hardware) : Les expérimentations ont été limitées par les capacités de calcul disponibles. Une puissance de calcul supérieure aurait permis de tester des résolutions plus hautes ou des itérations plus nombreuses (Sampling steps) qui influencent aussi le style.

Le "Prompt Engineering" : Malgré la rigueur des tests sur le CFG ou le Seed, le langage humain (le prompt) même si je l'ai généré avec une IA aussi (chatgpt) reste une barrière. L'interprétation des mots par l'IA peut parfois biaiser l'analyse d'un paramètre technique.

2. Les Perspectives

Cependant, ces travaux ouvrent des perspectives stimulantes pour la pratique architecturale :

Intégration à la maquette numérique : Une perspective majeure serait d'étudier comment ces paramètres de style peuvent être couplés directement à des logiciels comme Revit ou ArchiCAD, pour passer instantanément de la maquette technique au rendu d'ambiance maîtrisé.

L'IA comme outil de concertation : Explorer comment la rapidité de génération de styles différents peut aider les architectes à présenter plusieurs intentions architecturales à un client ou à des habitants lors de phases de concertation.

Éthique et droit d'auteur : La question de la "place de l'auteur" pourrait faire l'objet d'une recherche approfondie sur la propriété intellectuelle des styles générés : à qui appartient le style si l'IA a été entraînée sur des milliers d'images d'autres architectes ?

Vers une IA "spatiale" : Au-delà de l'image en 2D, la perspective suivante serait l'application de ces paramètres de contrôle à la génération de volumes 3D et d'espaces immersifs en temps réel.

Conclusion générale

Le questionnement initial de ce mémoire portait sur la capacité de l'architecte à contrôler le style graphique de ses rendus face à l'intelligence artificielle. À travers les expériences établies avec Stable Diffusion en mode Image-to-Image, j'ai cherché à comprendre jusqu'où l'utilisateur contrôle le résultat et reste maître de sa production.

Mes recherches démontrent que le contrôle du style par l'architecte n'est pas vraiment une illusion, mais il ne se situe pas là où je l'attendais. Par exemple le Prompt me semblait être au premier abord l'outil principal, néanmoins l'expérimentation prouve que ce n'est pas le cas, en réalité c'est l'Image de Référence avec un Denoising Strength équilibré qui définit le style.

L'autonomie de l'algorithme est dirigée par l'image de base, on peut dire donc (à travers l'expérimentation sur le modèle de stable diffusion) que l'IA ne remplace pas l'architecte, elle "interprète" sa base géométrique et apporte des changements en fonction des données entrées. La maîtrise du style est donc entre les mains de l'architecte en maîtrisant les différents paramètres.

Connaissance apportée

La connaissance apportée par ce travail réside dans la manipulation de l'outil stable diffusion, de comprendre ou aller directement et quels sont les paramètres à privilégier dans le cas d'une intention de modifier un style graphique d'une image architecturale,

La connaissance principale c'est que (dans stable diffusion) contrairement à ce que tout le monde pense, pour modifier le style graphique, ce n'est pas le prompt qu'il faut modifier et refaire à chaque fois, c'est plutôt l'image de référence en priorité, puis la gestion du denoising en fonction de la quantité de détails qu'on veut garder, ainsi que la compréhension des autres paramètres.

En conclusion, l'IA ne signe pas la fin de l'auteur. Elle transforme l'architecte en un pilote de paramètres. Le style graphique n'est plus seulement une affaire de coup de crayon ou de savoir-faire sur un logiciel de 3D classique, mais une compétence nouvelle liée à la gestion numérique et paramétrique, (l'image devient un modèle paramétré (figure)). L'architecte de demain sera celui qui saura doser l'autonomie de la machine pour servir sa propre vision créative.

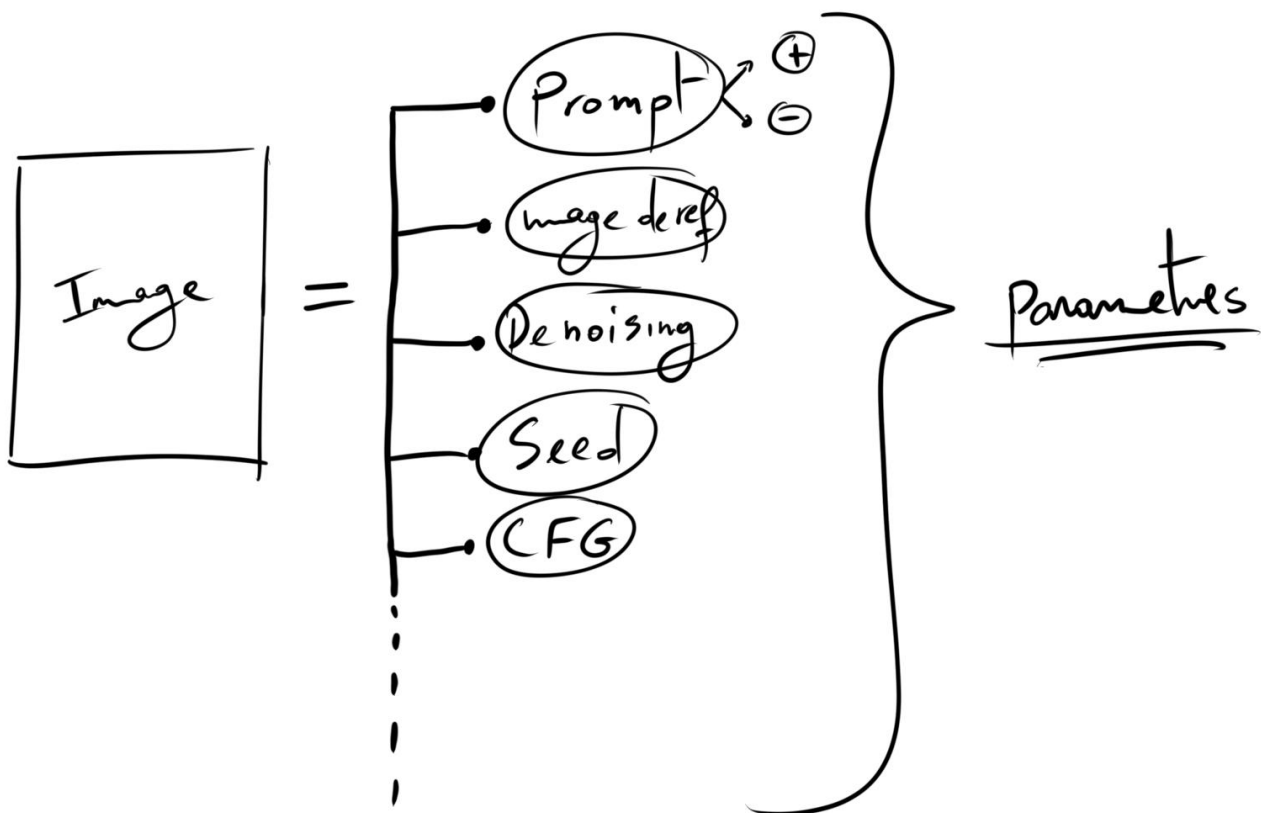


FIGURE 46 : L'IMAGE DEVIENT UN ENSEMBLE DE PARAMETRES

SOURCE : schéma personnel

RÉFÉRENCES

Articles

- ZHANG, Xin & LIU, Wenwen. "Boosting Architectural Generation via Prompts Report." arXiv, 24 avril 2024.
- DECKERS, J. et al. "Manipulating Embeddings of Stable Diffusion Prompts." 2024.
- ROMBACH, Robin, BLATTMANN, Andreas, LORENZ, Dominik, ESSER, Patrick & OMMER, Björn. "High-Resolution Image Synthesis with Latent Diffusion Models." 2022.
- TUMANYAN, V. et al. "More Control For Free! Image Synthesis with Semantic Diffusion Guidance." 2022.
- ISOLA, P., ZHU, J.-Y., et al. "Image-to-Image Translation with Conditional Adversarial Networks Pix2Pix." 2017.
- ZHU, J.-Y., PARK, T., et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks CycleGAN." 2017.
- MARSAULT, Xavier & NGUYEN, Hong Minh-Chau. "Les GANs stimulateurs de créativité en phase d'idéation." SHS Web of Conferences, 2022.
- KUPPENS, R. "Rôle des artefacts dans le projet d'architecture : qualification de représentations dans le cadre d'un concours." 2020.
- BEYAN, Eleonora Vilgia Putri & ROSSY, Anastasya Gisela Cinintya. "A Review of AI Image Generator Influences, Challenges, and Future Prospects for Architectural Field." 2023.
- LI, Yuqian, XU, Weiguo & LIU, Xingchen. "Research on Architectural Generation Design of Specific Architects' Sketch Based on Image-To-Image Translation." In Hybrid Intelligence, Springer Singapore, 2023.
- BARUA, Hrishav Bakul, KRISHNASAMY, Ganesh, WONG, KokSheik, STEFANOV, Kalin & DHALL, Abhinav. "ArthDR-Net: Perceptually Realistic and Accurate HDR Content Creation." APSIPA ASC, IEEE, 31 octobre 2023.
- HUANG, Jeffrey, JOHANES, Mikhael, KIM, Frederick Chando, DOUMPIOTI, Christina & HOLZ, Georg-Christoph. "On GANs, NLP and Architecture: Combining Human and Machine Intelligences for the Generation and Evaluation of Meaningful Designs." Technology|Architecture + Design, 5(2), 2021.
- GOODFELLOW, Ian, POUGET-ABADIE, Jean, MIRZA, Mehdi, et al. "Generative adversarial nets." Advances in Neural Information Processing Systems 27, Curran Associates, 2014.
- YAN, Xinkai, XU, Jieting, HUO, Yuchi & BAO, Hujun. "Neural Rendering and Its Hardware Acceleration: A Review." arXiv, 6 janvier 2024.

Science étonnante. "Stable Diffusion : IA génératives et méthodes de diffusion." 13 janvier 2023.

Sanz, F. "How Stable Diffusion works." novembre 2023.

[10 Visualization styles used in 3D architecture design [en ligne]. [consulté le 6 janvier 2025]. Disponible à l'adresse : <http://www.proglobalbusinesssolutions.com/architectural-visualization-styles/>

Mémoires et thèses

[EMMA, Kaidi. Les images de synthèse dans le but de convaincre ou faire comprendre le projet [en ligne]. Mémoire : École Nationale Supérieure d'Architecture de Paris-La Villette, 2019-2021. [consulté le 6 janvier 2025]. Disponible à l'adresse : https://atelier.dnarchi.fr/pluginfile.php/612/mod_folder/content/0/Kaidi_2021.pdf?forcedownload=1

[JULIEN, Faure. La communication, valeur fondamentale de l'architecte maître d'œuvre 2017[en ligne]. Mémoire : École Nationale Supérieure d'Architecture de Grenoble. [consulté le 6 janvier 2025]. Disponible à l'adresse : https://issuu.com/julienfaure/docs/ensag.hmonp.me_moire_en_quoi_la_co

Sites internet

[ARCHIGRIND, 2017] 7 Styles de Rendus d'Architecture [en ligne]. 23 novembre 2017. [consulté le 6 janvier 2025]. Disponible à l'adresse : <https://archigrind.fr/styles-de-rendu-darchitecture/>