Deep Learning : Les algorithmes de génération d'art visuel selon les différentes visions de l'art

Mémoire présenté par :

Helena Li

Pour l'obtention du M1 Miage

De l'université Panthéon Sorbonne

Année universitaire : 2022-2023

Date de soutenance : 04/07/2023

Directeur de mémoire : Camille SALINESI

Membre du jury: Rebecca DENECKERE

Table des matières

Remerciements

Je voudrais remercier dans un premier temps mon tuteur enseignant, Mr. Camille SALINESI, qui m'a beaucoup appris au cours de ce mémoire. En effet, il m'a guidé sur la méthodologie de recherche, m'a dirigé sur les problématiques potentielles ; et en somme il a été très pédagogue avec moi. Je souhaite donc le remercier pour ces raisons.

Je souhaite également remercier l'équipe pédagogique de l'université Paris 1 Panthéon Sorbonne pour la qualité du cursus, et pour tout ce que j'ai pu apprendre au cours de cette année.

Enfin, je remercie BNP Paribas - BCEF IT pour m'avoir accueillie en tant qu'alternante. Les équipes cybersécurité, et particulièrement l'équipe RCR 004 ont veillé à mon intégration et ont fait de mon travail une expérience plaisante. En particulier, je souhaite remercier mon maître d'apprentissage Elyes IGHILAZA pour sa pédagogie, sa compréhension, et sa bienveillance à mon égard.

Résumé

À la suite de la récente introduction et popularité de générateurs d'art au grand public, on pourrait se demander quelles sont les différents types de générateurs existants, comment ils fonctionnent et pourquoi ont-ils été codés ainsi ? Cet état de l'art vise à explorer les différents types de modèles de Deep Learning utilisés pour créer de l'art visuel, tel que des peintures, des croquis, ou des dessins digitaux. En effet, chaque personne a souvent une vision de l'art qui lui est propre, se traduisant par un objectif de génération d'image différent, et c'est ce à quoi on s'intéresse et développe dans cet état de l'art.

Mots clés : Deep Learning, Génération d'Art, Réseaux neuronales, Art, Peinture, Image, GAN, Décodeur, Encodeur

Glossaire

Acronymes	Définition
IA	Intelligence Artificielle
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
ANN	Artificial Neural Network
NLP	Natural Language Processing
CLIP	Contrastive Language Image Pre-training
LSTM	Long Short Term Memory

1. Introduction

1.1. Motivation et Contexte

1.1.1. Les origines du Deep Learning et impact sur la génération d'art

Le deep learning ou apprentissage profond est une famille de techniques faisant partie du domaine de l'apprentissage. Sa caractéristique principale est l'usage de réseaux de neurones artificiels afin d'effectuer l'apprentissage, qu'il soit supervisé, semi-supervisé, ou pas supervisé. Superviser un apprentissage en deep learning signifie qu'un humain contribue dans celui-ci, que ça soit en paramétrant l'algorithme ou en lui donnant des requêtes. Le premier article scientifique et algorithme fonctionnel de deep learning est attribué à Alexey Ivakhnenko et Lapa en 1967, qui utilisait un apprentissage supervisé avec plusieurs couches de perceptrons. [17]

L'utilisation de deep learning dans des industries s'est popularisée dans les années 2000 selon Yann LeCun, un éminent chercheur en intelligence artificielle et vision artificielle. C'est cependant en 2012 qu'une révolution du deep learning s'est passée, se concentrant sur les thèmes de la reconnaissance d'image et d'objet. Un exemple de cette révolution est la victoire d'AlexNet. Ce dernier est un algorithme de reconnaissance d'image utilisé sur ImageNet, un dataset d'images contenant plus de 14 millions d'images, divisé en plus de 20 000 catégories. AlexNet a été le premier algorithme réussissant à obtenir un taux d'erreur de 15.3% [18] sur la classification de ces derniers, gagnant la compétition de 2012.

En 2019, Yann LeCun, Yoshua Bengio et Geoffrey Hinton ont été récompensés par le Turing Award pour leurs travaux en deep learning, considéré maintenant comme une composante critique de l'informatique.

Mais si le deep learning était utilisé par l'industrie, le terme était encore méconnu par le grand public et son utilisation n'était pas orienté pour la génération d'art. C'est en 2018 qu'un premier site grand public de génération d'art a été mis en ligne :

Artbreeder[1]. Artbreeder permet à ses utilisateurs d'effectuer des mélanges d'images, chaque image parent possède des « gènes », des caractéristiques paramétrables, et l'image générée est considérée comme « enfant ». Artbreeder utilise StyleGAN, un modèle créé par Nvidia [20], utilisé également par « This Person Does Not Exist[2] ».

C'est en 2021 que DALL-E[3], produit par OpenAI, un laboratoire de recherche dirigé par Joshua Bengio, popularise les IA [4]génératrices d'image. DALL-E utilise à la fois la technologie NLP [5]et CLIP[6], permettant de « traduire » la requête d'utilisateur en une image, et le Diffusion Model pour la génération d'images. Ces derniers sont des chaînes de Markov utilisées pour l'entraînement en deep learning. [19]

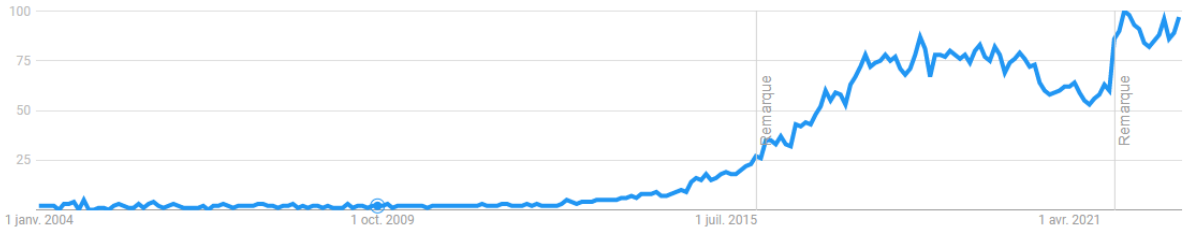

Figure 1 – Graphique représentant le nombre de recherche pour « Deep Learning » – Source : Google trends, 2004-2023

AlexNet a provoqué un pic de recherche sur Google en 2015, comme on peut l'apercevoir sur la figure 1, puis la sortie de DALL-E en essai public a provoqué un second pic en 2021, qui continue d'accroître encore aujourd'hui.

[1] https://www.artbreeder.com
[2] https://this-person-does-not-exist.com/en
[3] https://openai.com/product/dall-e-2
[4] IA : Intelligence Artificielle
[5] NLP : Natural Language Processing
[6] CLIP : Contrastive Language Image Pre-training

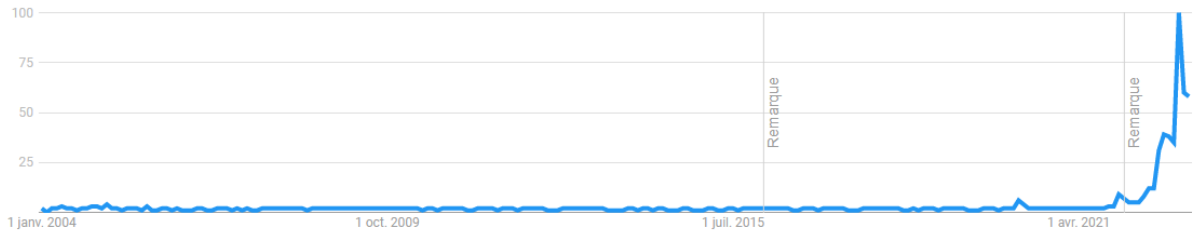

Figure 2 – Graphique représentant le nombre de recherche pour « art generation», – Source : Google trends, 2004-2023

De même, DALL-E a forgé l'intérêt du grand public pour l'art généré par des IA [figure 2]. S'ensuit alors de multiples projets de génération d'art : Midjourney, DALL-E 2, Nightcafe, hotpotAI, OpenArt…

1.1.2. Les différentes visions de l'art

Il n'existe pas une seule définition de l'art, bien au contraire, de nombreux philosophes et artistes possèdent leur propre définition de ce qu'est de l'art, ou non. Etymologiquement, le mot "art" est issu du latin "ars, artis", signifiant habileté, technique. Selon cette définition, et dans le cas où la conception d'une image par une machine est acceptée, les IA inspirées par cette vision seront orientées vers l'apprentissage de la technique : un coup de pinceau, pas de supervision humaine…

Pour d'autres, la définition de l'art est intrinsèquement liée au beau, à l'esthétique. Cette définition peut alors inclure des œuvres générées par une machine, comme DALL-E, car la notion de beauté est subjective. Une personne peut trouver qu'une œuvre de DALL-E est belle, et donc la considérer comme de l'art. De nombreuses IA génératives d'image permettent de créer des œuvres qui plaisent au grand public, que ce soient des œuvres abstraites, stylisées, ou réalistes.

Si l'acceptation ou non des œuvres générées par des machines en tant qu'art est encore débattu, il existe un rapprochement avec des mouvements comme le "Ready-Made", qui cherche à dépasser les définitions classiques de l'art. Les œuvres d'Andy Warhol rappellent également cette utilisation de la machine, usant l'imprimerie et la technique du silk-screening, par l'humain pour l'art.

En conclusion, la notion d'art et d'esthétique est différente pour chaque individu et de ce fait, les algorithmes mises au point pour générer de l'art vont également différer par leur objectif. Ce mémoire d'état de l'art cherche à explorer les différentes techniques utilisées pour générer de l'art, selon des visions de l'art différentes.

1.1.3. Les impacts et enjeux du Deep Learning et l'automatisation de l'art

L'introduction d'intelligence artificielle générative d'art implique de nombreux enjeux. Dans un premier temps, certains artistes ou industries utilisent cette technologie comme un outil, notamment pour de l'exploration artistique [2]. Les images générées, comme celles par le Colab Notebook, sont alors utilisées pour des publications sur des réseaux sociaux [12] ou bien des projets artistiques [13]. Dans un second temps, les IA génératives provoquent également des débats dû à leur capacité à remplacer un artiste humain, comme lorsqu'un utilisateur de Midjourney a utilisé l'outil pour participer à un concours artistique [21], dans la catégorie art digital. La victoire de l'utilisateur élève des inquiétudes concernant le remplacement des artistes humains, de la qualification des images générées (est-ce considéré comme de l'art, ou non ?) mais aussi des problèmes de droits d'auteur.

En conséquence, des plaintes ont déjà été portées contre des IA génératrices d'art, un exemple récent est la plainte en janvier 2023 de Sarah Andersen, Kelly McKernan et Karla Ortiz qui ont porté plainte contre Stability AI, Midjourney et DeviantArt.

En effet, les IA utilisées sont souvent de type multimodal - elles prennent un input humain sous forme de requête texte, puis génèrent une ou plusieurs images. L'entraînement de ces IA se font, comme dans le cas de OpenAI, sur des images que l'on peut trouver sur internet (google image, des sites de publications d'art comme DeviantArt [6] …). Autrement dit, les IA génératrices d'art s'entraînent sur des œuvres humaines sans en avoir le droit, et elles peuvent également copier des styles de dessin ou peinture d'artiste sans leur permission.

Certains pays, comme la Chine, ont décidé de prendre en main la régulation de ces IA génératives. [15] En effet, à partir du 10 janvier 2023, les images générées par des IA

qui ne sont pas indiquées en tant que telles seront bannies. Par ailleurs, en 2019, la Chine a également évoqué des règles concernant l'interdiction de deep-fakes [15,16], que les IA génératrices d'art peuvent également créer bien que ça ne soit pas leur fonctionnalité principale.

En conclusion, les IA génératrices d'art ont des impacts socio-économiques forts, qu'on ne peut cependant pas encore mesurer totalement car l'utilisation en masse de ces dernières est récente.

2. Méthodologie de recherche de la littérature

Afin de chercher des articles scientifiques, j'ai utilisé l'outil Google Scholar [7]. Pour débuter mes recherches, le thème étant très vaste, j'ai d'abord utilisé des requêtes assez générales : « *state of art 'art generation'* » pour voir ce qui existe déjà, puis « *art generation* », « *image generation* ». A cause de la nature même d'un état de l'art, j'ai décidé de filtrer par date, de sorte à n'obtenir des articles sortis entre 2016 et 2022 afin de n'avoir que articles récents. Durant cette étape j'ai également écumé des articles qui n'étaient pas encore fiables (articles de type « preprint »), mais qui me permettait de découvrir et préciser mon thème pour trouver des mots clés.

Afin d'effectuer une recherche de littérature scientifique, un exemple de revue systématique proposé dans [33] est un processus de dix étapes. Le principe est de d'abord définir la question de recherche et du périmètre, qui permettent de trouver des requêtes et mots-clés à utiliser. Cela permet également de réfléchir aux articles souhaités : si le but est de trouver des articles avec des résultats ou des types d'articles (méthode d'expérimentation...) spécifiques. La prochaine étape correspond à la sélection des articles, en ayant effectuée une recherche « globale » au préalable, des critères d'inclusion et d'exclusion sont alors utilisés. Le but est de ne que garder les articles répondant à la question de recherche définie plus tôt.

[7] https://scholar.google.com

Les articles sont ensuite classifiés : dans [33] la méthode « keywording of abstracts » est utilisée, consistant à lire l'abstract des articles et en déterminer des mots-clés de concept, contexte de ces derniers. Cela permet d'effectuer un premier travail de classification, mais aussi de mieux comprendre le set d'articles ainsi que de découvrir les possibles liens entre ceux-ci.

La dernière étape est une étape de *mapping* dans [33]. En utilisant la classification de l'étape précédente et en utilisant des outils comme Excel, les articles sélectionnés sont *mappés,* sous forme de *bubble plot*. Cela permet de visualiser et analyser le set d'articles (Ex : Les articles les plus récents sont-ils d'une catégorie spécifique ? Si oui, pourquoi ?), qui permet d'effectue une revue de la recherche et identifier les écarts éventuels.

Dans le contexte du mémoire d'état de l'art, j'ai utilisé une méthode de recherche plus simple : la méthode "Snowball", ou "boule de neige" [31]. Cette méthode consiste à trouver un premier set d'articles en effectuant une recherche par mot-clé, par exemple sur Google Scholar, qui permet d'obtenir une diversité d'articles de plusieurs éditeurs. Le but du premier set est d'être divers [31], et dans le cas où le sujet est vaste et concerne beaucoup d'articles, il faut privilégier les articles dont les citations sont nombreuses. Ce premier set était composé des premiers articles issues des requêtes générales ci-dessus. Par la suite, j'ai appliqué la méthode "backward snowballing" [31], qui consiste à partir d'un article du premier set, de lire et chercher de nouveaux articles dans les références de ce dernier.

Remarquant que les modèles proposés étaient essentiellement du Deep Learning et au fur et à mesure, que mon thème se précisait sur les différentes visions de l'art causant des modèles différents. J'ai limité mon périmètre d'art à de l'art visuel, et je cherchais alors donc des articles concentrés sur la génération d'art de type peintures, dessins, images digitales. J'ai détaillé mes recherches avec les requêtes « *art generation creativity* », « *deep learning art generation* ». Cela m'a fait découvrir un second outil de recherche, openacess the CVF[8]. Ce site regroupe des versions

[8] https://openaccess.thecvf.com/menu

gratuites d'articles scientifiques issues des conférences *IEEE International Conference on Computer Vision* (IEEE ICCV), identiques aux originaux si ce n'est pour les filigranes.

IEEE ICCV étant une conférence très fiable[9]J'ai également utilisé les requêtes « *art generation* », « *painting generation* » et « *image generation* » pour trouver des articles sur cet outil. Ayant maintenant rassemblé 28 articles, j'ai effectué une première élimination de 7 articles pour source pas assez fiables, la majorité de ces derniers étant des « preprint ». [Annexe 1]

Les critères d'inclusion et exclusion que j'ai choisie sont :

- Article Scientifique en anglais
- Pas de mémoire
- Contient des expérimentations et/ou data sets, car je souhaiterai, en M2, effectuer une expérimentation scientifique
- Pas de sources non fiables dont les articles en preprint, donc pas encore vérifiés par des pairs, des articles issus de cabinet de consulting/entreprise, de conférence de rang inférieur à B

Par la suite, dans un processus simplifié de mapping, j'ai commencé à lire les articles restants en diagonale, en prenant quelques notes, pour conserver uniquement ceux qui m'intéressaient et qui semblaient nécessaires pour construire mon mémoire. Cela s'est traduit par un document de tableau de synthèse [annexe 2]. Ce tableau contient l'article, le ou les techniques utilisées (modèle, algorithme...), le data set, et l'expérimentation. C'est par ce travail que mon plan de mémoire s'est détaillé, et qui m'a permis de choisir mes articles finaux pour mon mémoire. Sur 28 articles au total, j'en ai gardé 11 pour construire le cœur de mon mémoire, qui sont :

1. Wang, Alexander, Mengye Ren, and Richard Zemel. "Sketchembednet: Learning novel concepts by imitating drawings." In *International Conference on Machine Learning*, pp. 10870-10881. PMLR, 2021

[9] Selon http://portal.core.edu.au/conf-ranks/ - Rang A*

2. Colton, Simon, Amy Smith, Sebastian Berns, Ryan Murdock, and Michael Cook. "Generative search engines: Initial experiments." In *Proceedings of the International Conference on Computational Creativity*. 2021 == Source non-fiable mais potentiel pour l'introduction (les différentes utilisations d'art généré par des IA)
3. Bai, Zechen, Yuta Nakashima, and Noa Garcia. "Explain me the painting: Multi-topic knowledgeable art description generation." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5422-5432. 2021.
4. Lee, Byunghwee, Daniel Kim, Seunghye Sun, Hawoong Jeong, and Juyong Park. "Heterogeneity in chromatic distance in images and characterization of massive painting data set." *PLoS One* 13, no. 9 (2018): e0204430.
5. Xue, Alice. "End-to-end Chinese landscape painting creation using generative adversarial networks." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3863-3871. 2021.
6. Li, Zhuwen, Qifeng Chen, and Vladlen Koltun. "Interactive image segmentation with latent diversity." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 577-585. 2018.
7. Johnson, Justin, Agrim Gupta, and Li Fei-Fei. "Image generation from scene graphs." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1219-1228. 2018.
8. Liu, Songhua, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. "Paint transformer: Feed forward neural painting with stroke prediction." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6598-6607. 2021.
9. Park, C. and Lee, I.K., 2020. Emotional Landscape Image Generation Using Generative Adversarial Networks. In *Proceedings of the Asian Conference on Computer Vision*.
10. Tan, Wei Ren, Chee Seng Chan, Hernán E. Aguirre, and Kiyoshi Tanaka. "ArtGAN: Artwork synthesis with conditional categorical GANs." In 2017 IEEE International Conference on Image Processing (ICIP), pp. 3760-3764. IEEE, 2017.

11. Shahriar, Sakib. "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." *Displays* (2022) : 102237.

Le développement se compose en 3 parties :

1. Des synthèses d'articles, ayant pour but d'introduire des algorithmes et modèles.
2. Les techniques utilisées pour générer de l'art, en procédant naturellement à comment une machine décrit une œuvre, aux différents algorithmes principaux utilisés, et enfin des techniques de "surcouche" ou de perfectionnement.
3. La dernière partie porte sur l'expérimentation, d'abord sur les différents types de protocole de test qui ont été effectués dans les articles centraux du mémoire, puis une deuxième partie sur les data sets afin de remettre en contexte les tests. Enfin, cette partie se conclura sur une synthèse des résultats des articles centraux.

3. Résumés d'articles choisis

3.1. Synthèse 1 : GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network

Référence complète :

> Shahriar, Sakib. "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." *Displays (2022)*: 102237.

Dans cet article, on explore les différents modèles d'algorithmes de génération d'art basés sur les réseaux neuronaux artificiel de type GAN[10]. Il est traité ici de plusieurs formes d'arts : texte (roman, poésie), audio, visuel.

Le GAN est un type de modèle de deep learning avec deux acteurs principaux : un générateur, et un discriminateur. Fonctionnant sur le même principe que l'algorithme minimax, le générateur va générer un texte, une image, un élément. Cet élément généré sera envoyé au discriminateur avec des éléments réels du data set ; le déterminateur doit trier si l'élément qu'il a reçu provient du data set originel ou est un élément généré. Lorsque le discriminateur n'arrive plus, ou atteint un taux minimal de succès, cela signifie que l'algorithme a terminé son entraînement.

Ce modèle est très utilisé dans la génération, en combinaison avec d'autres types de réseaux neuronaux. Dans [11] trois types d'architecture de GAN sont présentés :

- Conditional GAN : Une extension du GAN basique, il permet d'ajouter des modules auxiliaires par exemple de classifications, ou d'input. Ce modèle est utilisé pour des générateurs multimodaux
- Deep Convolutional GAN / DCGAN : Des modèles utilisant le GAN et des CNN[11]. Ces derniers ont la particularité d'être performants pour déterminer des spatialités, très utilisés pour la génération d'art visuel.

[10] Generative Adversarial Network : voir 4.2.5
[11] Convolutional Neural Network : voir 4.2.2

- Recurrent Adversarial Network : Des modèles utilisant le GAN et des RNN[12], qui sont performants pour des tâches ayant une notion de séquentialité et de temporalité. De ce fait, ce type d'architecture est plus utilisé dans de la génération de texte ou d'audio.

Parmi les exemples de modèles, il y a un générateur d'art non conventionnel [12], montré dans la figure 3, utilisant le DCGAN et le data set WikiArt. Ce dernier contient 80 000 peintures de 1119 artistes différents. Le protocole expérimental utilisé est une constitution de deux sondages : le premier correspond à répondre oui ou non à la question "est-ce que cette œuvre a été faite par un humain ?" et une note de préférence allant de 1 à 5. Le résultat est que 53% des œuvres générées ont été classifiées comme œuvre humaine, et que les individus (n=18 participants) n'ont pas eu de préférence particulière entre les œuvres générées et les œuvres humaines (3.2 de moyenne contre 3.1 pour les œuvres humaines).

Figure 3 – Images générées avec le générateur d'art non conventionnel – Source : [12]

[12] Recurrent Neural Network : voir 4.2.3

3.2. Synthèse 2 : End-to-end Chinese landscape painting creation using generative adversarial networks

Référence complète :

> Xue, Alice. "End-to-end Chinese landscape painting creation using generative adversarial networks." In *Proceedings of the IEEE/CVF Winter Conference on* Applications of Computer Vision, pp. 3863-3871. 2021

Selon les différentes visions de l'art, des algorithmes différents sont créés. Dans cet article, des critiques sont émises sur les algorithmes actuels et sur le fait qu'ils ne sont pas originaux dû à leur dépendance à la supervision. De plus, de nombreux algorithmes sont axés uniquement sur la génération d'art occidental.

Dans [5] le modèle proposé est celui de figure 4, permettant de génération des peintures traditionnelles chinoises sans input humain :

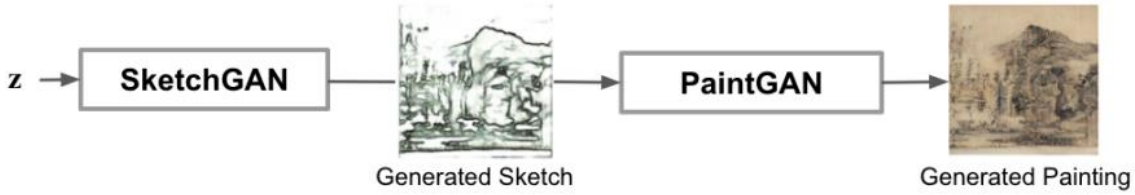

Figure 4 – Architecture du générateur d'art traditionnelle chinoise – Source : [5]

Comme montré dans la figure 4, SketchGAN est réseau neuronal de type "Generative Adversarial Network", ou GAN, entraîné sur des images correspondant à des esquisses de peinture. Ces esquisses sont elles-mêmes générées par des algorithmes détectant les "bords" des peintures issues du data set. Une fois l'esquisse générée, un second GAN intervient sous le nom de PaintGAN.

Ce dernier est entraîné sur les paires des réelles esquisses et des peintures comme indiqué sur la figure 5, et génère la peinture pour l'esquisse générée, permettant d'obtenir le résultat final. Le data set est basé sur 192 peintures traditionnelles chinoises, récupérées de plusieurs musées américains. Ces peintures ont été rognées de sorte à ne pas garder les parties contenant de la calligraphie.

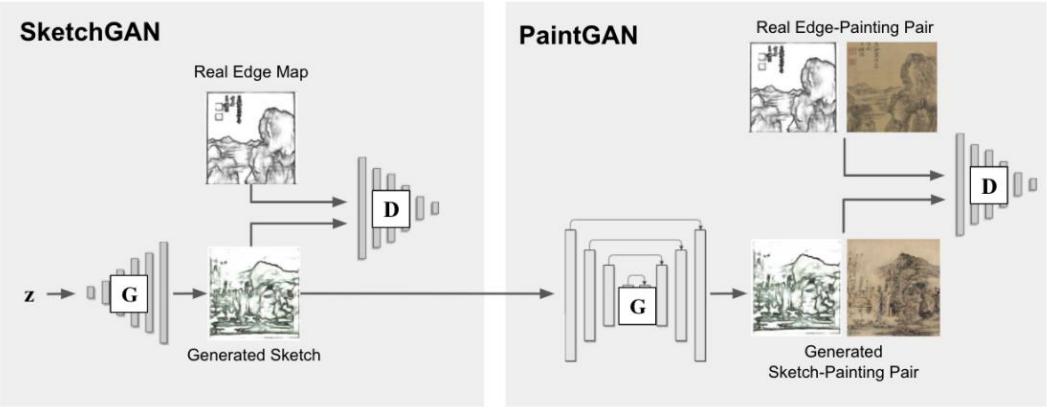

Figure 5 – Architecture interne de SketchGAN et PaintGAN – Source : [5]

Le protocole d'expérimentation de cet article réside sur deux sondages, sur un échantillon de 242 participants : l'un demande à des individus si l'œuvre devant lui a été peint par un homme ou par une machine, et une note allant de 1 à 10 sur la confiance de sa réponse. Le second sondage consiste à noter quatre critères : esthétique, clarté, composition et créativité. Pour chacun des critères, il existait quatre réponses possibles : désaccord, un peu en désaccord, un peu en accord, accord. Les résultats sont comparés à deux algorithmes Baseline : RaLSGAN et SAPGAN, dans une moyenne. Le résultat est que le modèle a été mieux noté sur les deux sondages.

3.3. Synthèse 3 : Paint transformer: Feed forward neural painting with stroke prediction.

Référence complète :

> Liu, Songhua, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. "Paint transformer: Feed forward neural painting with stroke prediction." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6598-6607. 2021

Dans [8] l'approche de l'art est plus humaine, spécifiquement axé sur la question : comment est-ce que l'on peint ? De la même manière que des artistes humains apprennent parfois la peinture via la copie, il s'agit ici d'un modèle d'algorithme permettant, à partir d'une photo ou d'une image, de déterminer les coups de pinceaux qui permettent de la peindre. Ce modèle est nommé "Paint Transformer", et est très différent de l'article précédent.

Dans la figure 6, on peut voir que l'on utilise le réseau neuronal CNN[13] dans [8] pour Paint Transformer, et en particulier le DETR[14]. En effet, les réseaux neuronaux de type CNN sont plus appropriés lorsqu'il y a besoin d'une notion de spatialité [11].

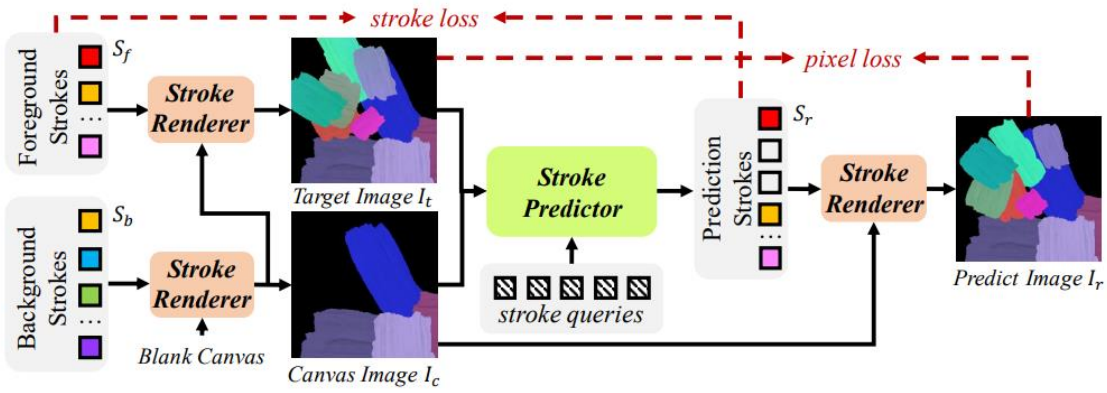

Figure 6 –Modèle de génération de coups de pinceaux – Source : [8]

Le fonctionnement de d'algorithme repose sur un *stroke renderer*, qui effectue un coup de pinceau, et un *stroke predictor*, qui prédit où ce coup et comment (taille, direction) il doit être. Le modèle s'entraîne sur lui-même : l'image It est une image avec des coups

[13] CNN : Convolutional Neural Network
[14] DETR Detection Transformer : https://arxiv.org/abs/2005.12872v3

de pinceaux générées aléatoirement. Le stroke predictor prédit alors le coup de pinceau dans l'image Ir, basé sur le canvas Ic également généré aléatoirement. Des comparaisons sont effectuées entre l'image target It, et le résultat Ir, permettant d'entraîner l'algorithme selon la précision de la prédiction.

Le protocole expérimental de cet algorithme consiste à des tests de performance, effectués avec une carte graphique Nvidia GTX 2080Ti, en comparaison avec deux autres algorithmes : RL et Optim. De la même manière, il y a également un test d'efficience et de précision (nombre de coups de pinceaux utilisés et position des coups de pinceaux) avec ces mêmes algorithmes. Le résultat est que Paint Transformer performe mieux que sa Baseline, et sur la comparaison d'efficience, Paint Transformer performe mieux sur 3 critères sur 5.

4. Techniques utilisées pour créer de l'art visuel

Nous allons d'abord voir le processus général de la génération d'image, dont nous préciseront ensuite les multiples méthodes possibles.

Dans un premier temps, nous verrons comment décrire une image. En effet, les générateurs d'image ont besoin de caractéristiques, d'attributs afin de s'entraîner puis produire des résultats. Par exemple, si l'on souhaite générer une peinture de chien, il faudrait déjà caractériser ce qu'est un chien, à quoi cela ressemble, qu'est-ce une peinture et qu'est-ce qui caractérise une peinture (ex : coups de pinceaux, certaines teintes de couleur…).

Dans un second temps, nous allons voir les différents modèles utilisés pour générer des images, leur architecture générale, leur fonctionnement et dans quels types de générations ils sont avantageux.

Dans un dernier temps, nous verrons l'utilisation et les choix de modèle de générateurs selon des buts et vision artistique différents. Ces visions artistiques différentes incluent parfois l'utilisation d'autres technologies en plus de la génération d'image, ou poussent parfois à utiliser plusieurs modèles de générateurs en même temps.

4.1. Classification, segmentation : comment décrire une image ?

Afin de générer de l'art visuel, les IA, dans le contexte d'utilisation de deep learning, s'entraînent sur des data sets. Ces derniers contiennent les données, essentiellement des images, qui permettent un premier apprentissage. Nous cherchons alors à lister les types de données utilisées : comment est-ce qu'une IA comprend, ou décrit une image ?

Une image correspond à un fichier, le plus souvent de format jpeg ou png. La différence entre des types de fichiers est que le format png est de qualité plus haute. Cette qualité est jaugée par le pixel, qui correspond à l'unité la plus minimale d'une image numérique et est caractérisée par une couleur primaire (RVG – Rouge, Bleu, Vert) et supporte la transparence (bien que ce dernier point ne soit pas pris en compte lors de l'entraînement d'IA génératrice d'art). Plusieurs caractéristiques peuvent en être extraites d'une image :

- La segmentation
- Des classifications

4.1.1. La segmentation

La segmentation correspond au découpage d'une image en plusieurs éléments comme dans [6]. Traditionnellement en dessin, ce découpage permet de diviser une image en zones d'ombres et de lumières, appelés valeurs. En génération d'image, on essayera plutôt de distinguer les composantes d'une image ; comme distinguer un personnage de son décor. L'objectif reste cependant similaire : le but est de dégager la composition d'une œuvre. La segmentation est d'autant plus importante dans des modèles de génération multimodales, car l'IA doit comprendre à quoi correspondent les différents mots dans une requête, et les assembler dans une image.

Selon [6], la segmentation peut être vue comme un problème « few shots active learning ». « Few shots » correspond à la capacité, pour une IA, d'apprendre de nouveaux concepts rapidement et avec un data set à la quantité limité, tandis qu'« active learning » correspond à l'idée qu'un échantillonnage intelligence de data sets permet de créer des modèles plus performants et précis.

Dans [6] on utilise un classifieur qui possède deux labels : arrière-plan et premier-plan, ce dernier doit pouvoir classifier tous les pixels d'une image en un des deux labels. Selon [23], les CNN permettent d'obtenir les meilleures performances concernant ce type de tâches maintenant.

Figure 7 – Exemple de segmentation à partir d'une photo – Source : [6]

Dans [5], on utilise également la segmentation : d'une part, pour trouver les contours d'une image et générer des esquisses (SketchGAN), puis pour peindre (PaintGAN).

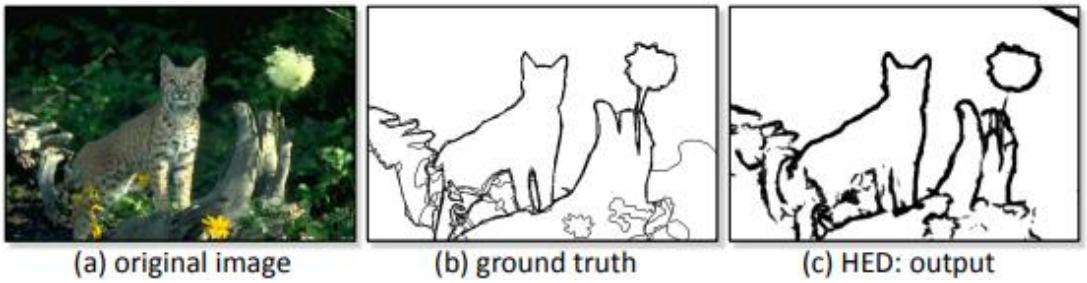

Figure 8 – Exemple de segmentation de contour avec HED, utilisé dans SketchGAN – Source : [24]

Selon [24] les CNN obtiennent de meilleures performances pour la détection des bords d'un dessin, comme on peut le comparer entre la figure 8, qui utilise un CNN, et la figure 9 qui n'utilise pas de réseaux de neurones artificiels. Le (b) ground truth correspond au résultat recherché ou attendu, tandis que (c) correspond au résultat obtenu.

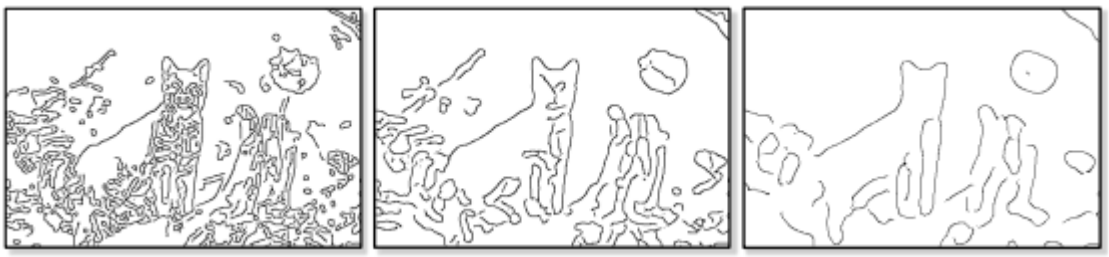

Figure 9 – Exemple de segmentation par Canny Detector – Source [24]

Le principe ici est de hiérarchiser les éléments d'une image afin d'en déterminer les contours, d'où l'utilisation de CNN qui comme dit dans [8,11] est plus utilisé lorsqu'il y a des notions de spatialité.

4.1.2. Les classifications

En plus des caractéristiques inhérentes de l'image, il existe des classifications plus ou moins subjectives que l'on peut préciser. Cela peut correspondre au style et au mouvement artistique auquel appartient une image, mais aussi des métriques spécialisées comme dans [4,9]. Les classifications constituent une partie intégrante aux IA génératrices multimodales [10,11] mais peuvent également être utilisées pour s'entraîner sur de larges data sets.

Ces classifications peuvent faire partie du data set, comme c'est le cas avec ImageNet, qui contient des annotations pour chaque image. Cependant, nous pouvons également obtenir des classifications par l'image elle-même comme c'est le cas dans [4]. La métrique utilisée est celle de la densité de couleur par pixel, car chaque artiste, chaque mouvement utilise les couleurs différemment. On peut apercevoir la différence dans la figure 10, comparant deux œuvres d'artistes aux styles très distincts. En observant la densité de couleur utilisée par pixel, cela permet de catégoriser une œuvre dans un mouvement, ou du moins d'obtenir un indicateur par groupe.

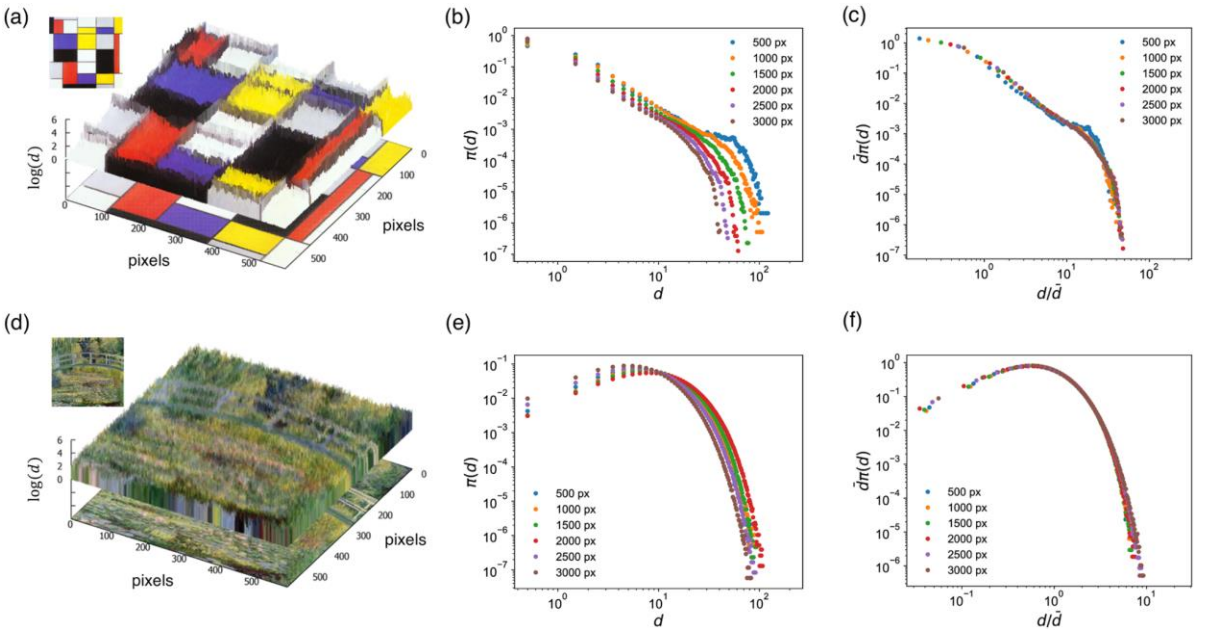

Figure 10 – Comparaison de densité de couleur par pixel, avec une œuvre de Piet Mondrian15 et une œuvre de Claude Monet16 – Source : [4]

[15] https://www.wikiart.org/en/piet-mondrian/composition-a-1923

[16] https://artmuseum.princeton.edu/collections/objects/31852

Dans [3], un modèle est proposé afin de générer des descriptions de peinture composées par trois éléments, montré dans l'architecture représenté dans la figure 11 :

- Le contenu : Une description claire de la peinture. Celle-ci précise les éléments de la peinture elle-même et non son contexte
- La forme : Une description stylistique de la peinture, et des émotions qu'elle peut évoquer
- Le contexte : Le contexte de la peinture, c'est-à-dire son époque, son auteur, la raison de la peinture etc. Ces informations sont extraites de Wikipedia puis resynthétisé

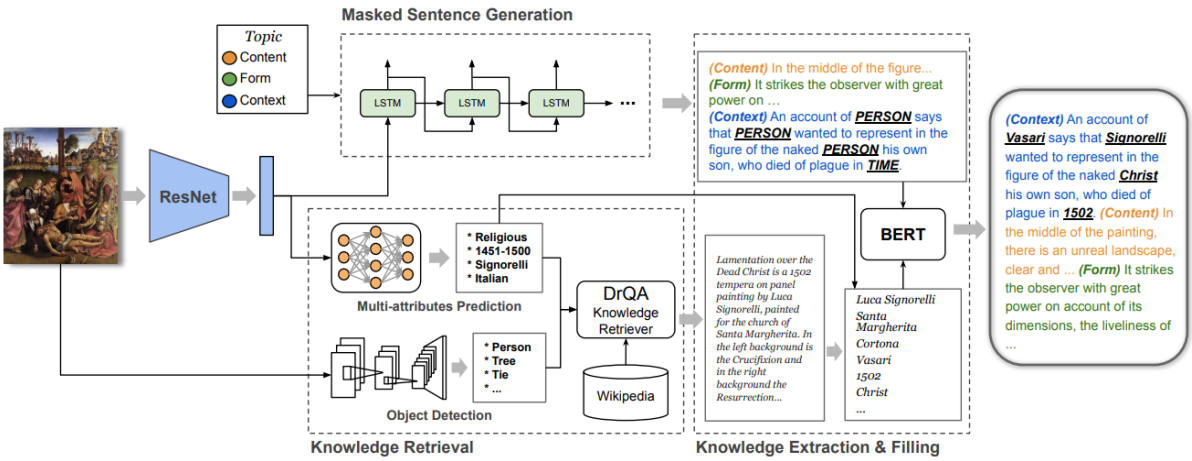

Figure 11 – Modèle de génération de descriptions à partir d'une peinture – Source : [3]

Selon [3], la classification la plus utilisée est celle du contenu, et dans de nombreux data sets comme Wiki Art [3], la classification est plutôt celle de forme + auteur. Cependant, pour [3] on ne peut réduire une peinture qu'à un seul mouvement ou une seule signification. L'art peut être un mélange de différentes idées et mouvements, et chaque peinture ne possède pas qu'un seul attribut distinctif, d'où l'importance d'avoir des descriptions multiples pour chaque œuvre. Ces descriptions sont générées à l'aide de ResNet, un réseau neuronal, afin d'extraire les informations de Wikipédia pour le contexte, puis remplir des structures grammaticales avec ces dernières.

4.2. Réseaux de neurones : les modèles d'algorithmes de Deep Learning

Les réseaux de neurones artificiels constituent une famille de techniques et modèles dans le domaine du deep learning. Ils se déclinent en plusieurs types de réseaux de neurones, mais comportent tous plusieurs couches de nœuds, qui représentent les « neurones » :

- Une couche d'entrée de données, « input layer », indiquée dans la figure 12 en vert
- Une couche de sortie de données, « output layer », qui contient le résultat, indiquée en jaune dans la figure 12
- Les couches intermédiaires « hidden layer » où se déroulent le ou les traitements des données d'entrées, comme indiqué dans la figure 12 en bleu.

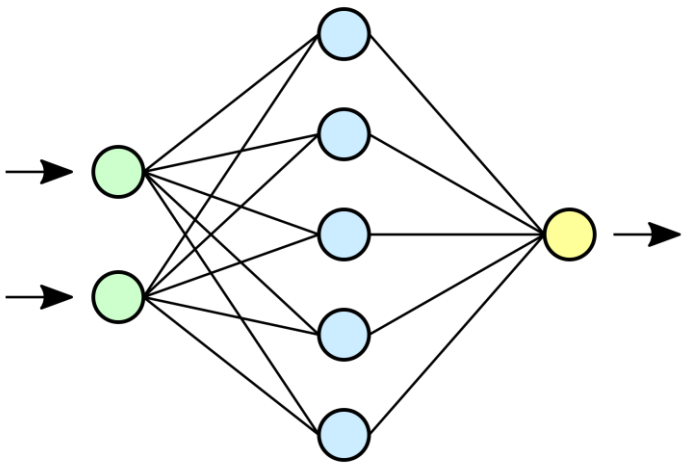

Figure 12 – Exemple de réseau de neurone simple – Source : [32]

Chaque type de réseaux de neurones possèdent ses caractéristiques, ses avantages et ses inconvénients.

4.2.1. Encodeur / Décodeur

Les termes encodeurs et décodeurs sont utilisés tout au long de ce mémoire, que nous expliciterons pour une meilleure compréhension des parties à venir :

- Encodeur : L'encodeur permet de convertir une donnée dans un format nécessaire. En génération d'image, cela peut être par exemple convertir un fichier d'image jpeg en des vecteurs ou matrices, dont chaque cellule représente un pixel de l'image.
- Décodeur : Le décodeur effectue l'inverse l'encodeur. Il effectue une transformation d'un format nécessaire, au format de l'output, soit le résultat. Par exemple, transformer des vecteurs en image.

Les termes encodeurs et décodeurs sont généraux, et un encodeur peut par exemple être un réseau de neurones comme c'est le cas dans [1] où l'encodeur est un CNN et le décodeur un RNN.

4.2.2. Convolutional Neural Network (CNN)

Les CNN sont un type de réseau de neurone artificiel dont la particularité est que la couche d'entrée, et donc les inputs, n'affectent pas tous les outputs, comme montré dans la figure 13. Autrement dit, contrairement à d'autres architectures, chaque neurone dans une couche CNN n'est pas connecté à tous les autres neurones de la prochaine couche. Cette flexibilité et capacité à travailler plus en détail les rendent excellents pour tout travail ayant une notion de spatialité [11].

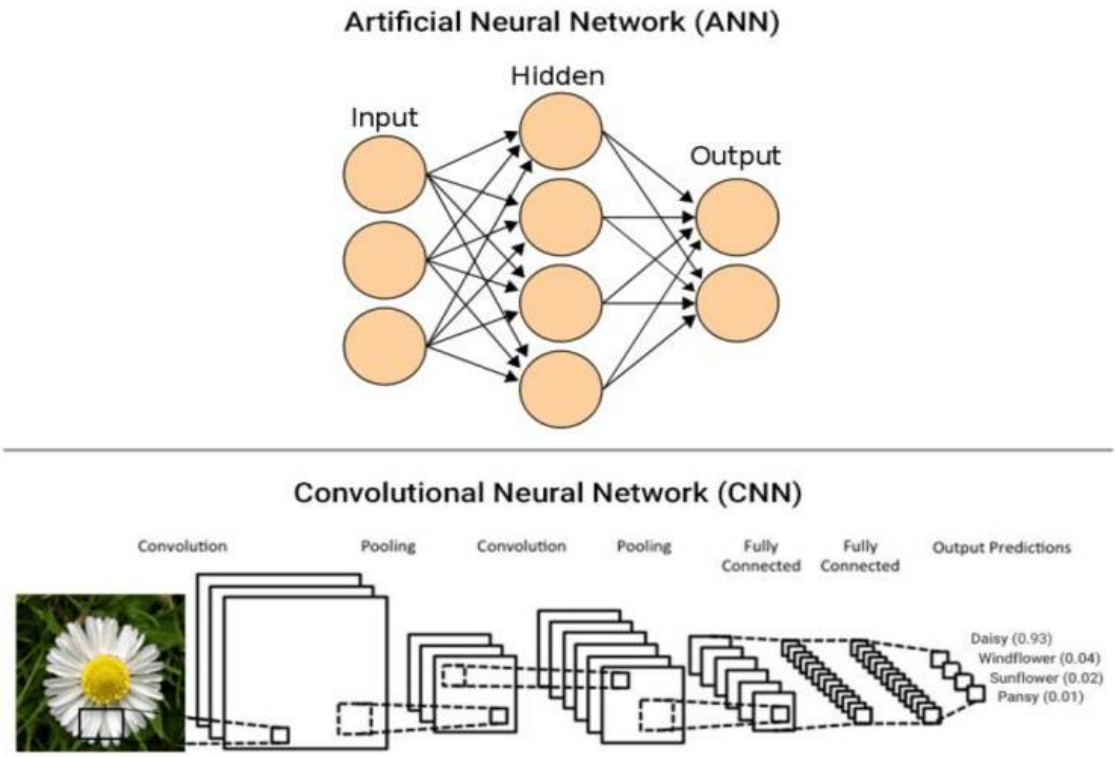

Figure 13 – Comparaison d'un ANN et CNN – Source : [27]

Ils sont principalement utilisés dans les IA génératives d'art, notamment avec des GAN que nous verrons ci-dessous, qui utilisent des CNN comme générateurs. En particulier, les CNN sont utilisés dans des cas où on a besoin de reconnaissance de *patterns*, qui peuvent être des bords ou des segmentations comme vu dans [5].

Les *features* correspondent aux patterns que l'on souhaite reconnaître. Ces features peuvent correspondre à des bords/contours, comme dans le SketchGAN dans [5], des segmentations d'image comme dans [6]. Dans ces CNN plus complexes, ces features peuvent être des classifications d'image, comme son genre, ou bien des descriptions de cette dernière avec les objets et entités composant l'image. Ces derniers sont plus utilisés dans des IA génératives d'art multimodales.

Le processus permettant de vérifier la présence d'une feature dans un input est appelé *filtering*. Une taille de fenêtre est paramétrée en pixel, correspondant à la surface de l'image qui sera examinée : pour chaque pixel correspondant à la ou les features, une valeur y est attribuée. Par exemple, si les pixels correspondent à la feature, un point

positif sera attribué, et dans le cas contraire ça sera un point négatif. Le résultat est une fenêtre de même taille, mais dont les valeurs sont la correspondance de la fenêtre et de la feature : on appelle ce résultat *feature map*.

Un CNN peut comporter plusieurs couches de filtering, selon son niveau de complexité. Ces couches sont nommées convolutions, comme indiqué dans la figure 12, car on va rechercher un pattern sur toutes les fenêtres possibles de l'input. Cela créer beaucoup de données, et c'est pour cela que la *pooling layer* permet de rétrécir le jeu de données en ne prenant que la donnée maximum dans une fenêtre de la *feature* map.

Ensuite, les données sont normalisées puis mises sous forme de liste : c'est la *fully connected layer*, qui nous permet d'obtenir le résultat final. Ce dernier est ensuite comparé aux inputs, et les paramètres sont modifiés afin de se rapprocher de l'input d'entrée : c'est la *rétro-propagation*. Cette dernière phase permet d'ajuster l'entraînement du réseau de neurones, et permet son apprentissage. Par exemple, la fully connected layer pourrait représenter des contours d'image sous forme de données (une valeur dans une matrice pouvant représenter la valeur d'un pixel), ou une classification d'image, qui seront comparés et vérifiés par les données de la couche d'entrée.

4.2.3. Recurrent Neural Network (RNN)

Les RNN sont un type de réseau de neurones artificiels mis au point en 1925 par Lenz et Ising. La caractéristique principale du RNN est que contrairement aux CNN qui se ne se propagent que dans un sens (entrée vers sortie), les couches cachées du RNN peuvent contenir de la récurrence comme montré dans la figure 14. Cette dernière s'exprime par le fait que dans un RNN, une sortie ou output peut être ré-utilisée comme entrée ou input dans la couche cachée.

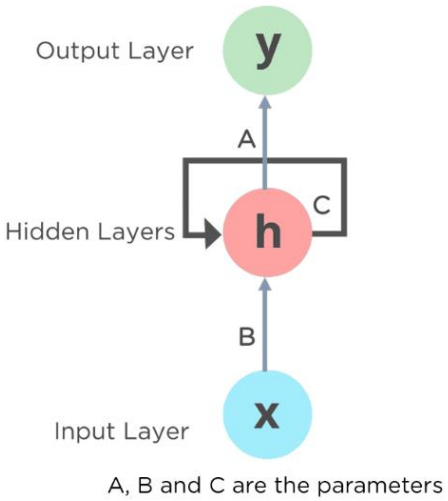

Figure 14 – Exemple de modèle RNN – Source : [30]

Les RNN, par leur nature récurrente, sont utilisées dans les domaines où la notion de temporalité et mémoire sont nécessaires, comme la musique ou le texte [11]. Il existe de nombreux sous-types en RNN, dont les architectures sont plus complexes et se composent de plusieurs couches de RNN, de différentes directions. En effet, si l'avantage des RNN est leur capacité de prédiction grâce à la récurrence, ils sont par conséquent pas très performants et rencontrent des problèmes de type *vanishing gradient* et *exploding gradient* [28,29] qui diminuent la précision du réseau de neurones en rendant l'apprentissage moins effectif.

Bien que l'utilisation de RNN soit plus rare en génération d'art visuel, il est utilisé dans [1] afin de déterminer dans quel ordre des traits ou coup de pinceaux sont faits ; dans

l'optique d'apprendre à une machine à dessiner ou peindre comme un humain, étape par étape.

4.2.4. Long Short-Term Memory (LSTM)

Les LSTM sont un type de RNN plus performant lorsqu'il y'a besoin d'une notion de "mémoire", et permet de résoudre certains problèmes des modèles RNN classiques, comme le *vanishing gradient* [28]. Les LSTM ont été mis au point par Hochreichter et Schmidhuber [29] en 1997. Ils sont composés de trois couches, comme indiqué à droite dans la figure 15 (a) ci-dessous, dont une d'entrée (input), sortie (output), et une couche récurrente LSTM.

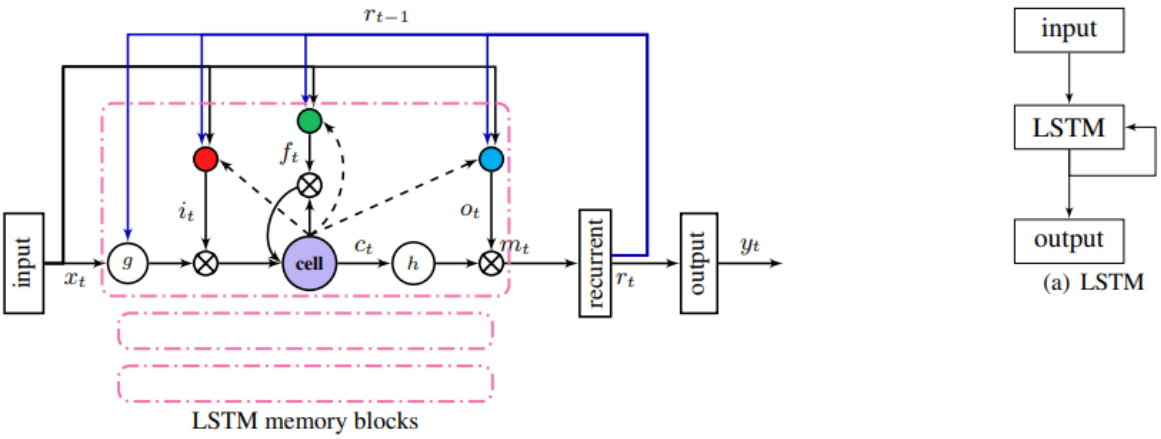

Figure 15 – Exemple de modèle LSTM et ses unités – source : [28]

La caractéristique des LSTM est qu'elle est composée de cellules possédant trois *gates*. Les *gates* permettent de définir la mémorisation d'un élément sous trois statuts : input, output, et forget. L'input permet de mettre à jour les éléments composant la "mémoire", comme l'occurrence d'un mot, l'output permet d'effectuer la prédiction du prochain mot, tandis que le forget gate permet de définir si un élément doit être porté à attention, ou peut être ignoré.

Par sa nature à pouvoir imiter une "mémoire", le LSTM est plus utilisé avec des formats de texte ou des mélodies. C'est notamment le cas dans [3] qui utilise le LSTM pour

générer des descriptions de peinture (*masked sentence generation*), ou dans [9] afin de classifier puis générer des paysages naturels émotionnels.

4.2.5. Generative Adversarial Network (GAN)

Le GAN est une famille de technologie utilisée en deep learning, généralement couplée avec des réseaux de neurones artificiels lorsque le but est de produire une œuvre artistique. Comme indiqué ci-dessous dans la figure 16, le GAN se compose de deux acteurs principaux :

- Le générateur, dont le but est de générer des résultats. Dans le contexte de la génération d'art visuel, ce sont des images générées à partir d'un data set ou input. Ces images sont alors envoyées au discriminateur et mélangées avec des celles issues du data set (représentées sur la figure sous le terme "real image")
- Le discriminateur a pour rôle de recevoir une image, réelle ou non, et d'en déterminer sa nature. Le discriminateur est en compétition avec le générateur : ce dernier cherche à générer des résultats qui se rapprochent du data set, tandis que le discriminateur doit pouvoir correctement différencier les résultats générés des images du data set.

Il peut exister plusieurs générateurs et discriminateurs au sein d'un même GAN, en sachant que chaque générateur ou discriminateur représente un réseau de neurones.

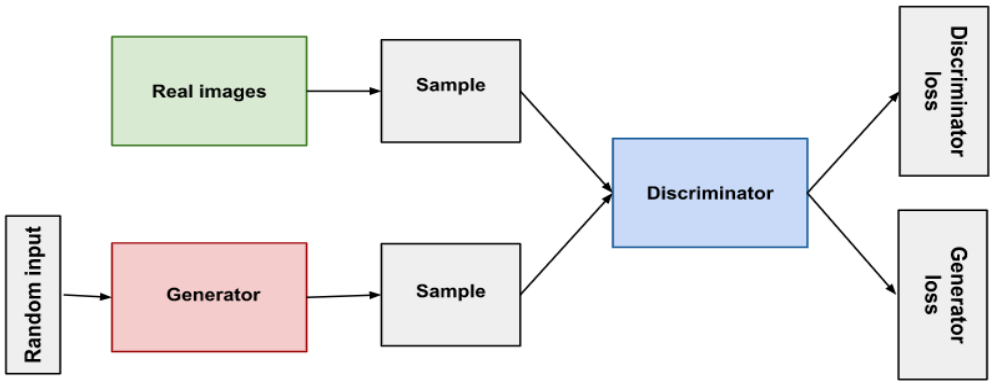

Figure 16 – Exemple de modèle GAN – Source : [33]

L'architecture décrite ci-dessus dans la figure 16 est inspirée du *minimax*, plus connue pour son utilisation pour créer des IA de jeux. Tout comme le *minimax*, le contexte du générateur et discriminateur est un jeu à somme nulle. Le générateur et discriminateur sont considérés comme des joueurs opposés dont le but est de minimiser leur perte. L'objectif final est que le discriminateur perde, signifiant que l'entraînement du générateur est devenu suffisant pour tromper le discriminateur, et donc que les images générées sont indiscernables de celles du data set.

En outre, le « random input » sur la figure 16 est souvent un bruit numérique [11]. Il correspond à une dégradation, une perte de qualité d'une image, se traduisant par des pixels parasites (pixel de couleur aléatoire, ou de couleur plus foncée ou claire) donnant un aspect granuleux à l'image. Le bruit numérique est utilisé durant l'entraînement afin de former des images composées de bruits aléatoires et les reconstituer jusqu'à ce qu'elles correspondent aux images d'entrée. D'autre part, une image peut être constituée de bruit afin de servir de schéma : les nuages de pixels colorés des bruits permettent de déterminer des compositions de façon simple, puis de faire générer des images de plus haute qualité par-dessus.

Afin de déterminer les pertes et victoires, et donc paramétrer l'entraînement du générateur, il existe une fonction « objectif » (loss function). Pour chaque algorithme utilisant le GAN, il existe une fonction objectif qui lui est propre. Le GAN est l'un des modèles les plus utilisés dans la génération d'image, en conjonction avec des CNN comme dans [2,5,7,9,10,11].

4.3. Multimodalité, ajout de style : différentes technologies pour différentes visions

Il existe plusieurs types de réseaux de neurones afin de produire des générateurs d'art différents, avec leurs propres caractéristiques. Selon l'objectif et le type d'œuvre visé, un ou des types de réseaux de neurones sont utilisés : la diversité des algorithmes est également dû à une diversité de la vision de l'art. Cette partie se consacre à détailler pourquoi et comment les réseaux de neurones ont été choisi selon l'objectif du générateur, ainsi que de montrer la vaste étendue d'utilisation des réseaux de neurones artificiels.

La multimodalité correspond aux générateurs d'art possédant plusieurs modes de génération, comme DALL-E : on génère de l'art visuel (image) par une requête (texte), d'où le terme « multimodalité ». Ce type de générateurs sont utilisés pour [2] :

- Faire de l'exploration artistique
- Générer des esquisses en grande quantité pour une industrie (ex. Industrie de la mode)

L'avantage des générateurs multimodaux est leur est de pouvoir effectuer une requête précise. Cette requête texte est ensuite traitée grâce au Natural Language Processing (NLP), un procédé similaire à la génération de description dans [3].

Un autre type de générateur est le modèle dans [5], où la vision de l'auteur considère que pour que l'on puisse considérer une image générée comme de l'art, il faudrait que le générateur ne soit pas supervisé. Le modèle proposé est un GAN qui est entraîné sur un data set précis afin d'apprendre à la machine à produire des œuvres d'un style précis, et ce sans qu'un humain intervienne dans le processus de création : c'est « l'inverse » des générateurs multimodaux. Cette approche concorde également avec [1,8], car bien qu'entraîné à reproduire symboliquement des images, le générateur ne prend pas d'autres inputs que celui du data set, ils ne sont pas guidés ni ajustés par des humains ensuite.

Cependant, une critique qui est émise serait que ces générateurs possèdent des biais [36] dû au fait qu'ils sont limités à des data sets, et donc ne peuvent que donc reproduire un style, un type d'œuvre sans dégager d'originalité.

Dans [36] ce problème est nommé « representional bias », un exemple donné est un générateur entraîné sur 45 000 peintures de la renaissance, notamment des portraits d'ethnicité européenne, et performe alors de manière insatisfaisante lorsque les requêtes dévient de cette catégorie.

Certains générateurs auront plutôt l'objectif de reproduire la technique humaine, et en quelque sorte, de faire apprendre à la machine à peindre comme un humain. Dans [8], le générateur a pour but d'apprendre l'ordre des coups de pinceaux dans une peinture, et d'appliquer ensuite cette technique selon une image donnée. La notion d'ordre de coups de pinceaux est rendue possible grâce aux RNN, et le résultat permet d'obtenir un effet de peinture traditionnelle, contrairement à beaucoup de générateurs qui ont plus un aspect lissé issu de l'art digital. En revanche, ce générateur ne permet que de reproduire une image.

Le modèle dans [1] utilise également le RNN ainsi que les CNN, et a un objectif similaire : il s'agit d'apprendre à représenter une image, coup par coup. Le résultat donne alors une représentation symbolique simplifiée d'une image, qui est très différent des résultats de [8] malgré une approche similaire. Cependant dans les deux cas, lorsqu'on souhaite un apprentissage de la technique humaine, étape par étape, il y a une préférence pour l'utilisation du RNN qui est plus adapté pour les notions de temporalité. Ces générateurs étant limité dans la créativité, ils peuvent être combinés avec d'autres générateurs comme une surcouche de « style » afin de donner un aspect plus traditionnel à une œuvre [8].

Pour d'autre, dans [9] le modèle proposé a pour but de générer des paysages naturels qui évoquent des émotions. Selon Paul Cézanne, « Une œuvre d'art qui n'a pas commencé dans l'émotion n'est pas de l'art. » [9] cherche à évoquer de l'émotion par l'image. Plutôt que de proposer un modèle de génération, c'est plutôt une nouvelle unité, le Emotional Residual Unit (ERU) qui est montré, ainsi qu'une fonction objectif. En utilisant le LSTM, le ERRU permet de classifier des images de paysage en termes d'émotions, puis de générer d'autres images par la suite avec un GAN. Cette vision pourrait permettre alors de générer des images spécifiquement selon une requête liée à l'émotion, plutôt qu'à des descriptions d'image.

Enfin, le tableau 1 ci-dessous synthétise les différents algorithmes utilisés par le set d'article, ainsi que leur objectif.

Tableau 1 - Synthèse des sources, objectif et algorithmes

Source	Modèle/algorithme	Objectif
1	CNN, RNN	Reproduire la technique humaine – représenter des images en symbole
2	GAN	Générer des œuvres par requête (multimodal)
3	LSTM, NLP	Générer des œuvres par requête
4	Metropolis-Hastings Process	Classification – Densité de couleur par pixel
5	GAN + CNN	Générateur sans input humain – Générer des peintures traditionnelles chinoises
6	GAN	Segmentation
7	GAN	Générer des œuvres par requête (multimodal)
8	CNN, RNN	Reproduire la technique humaine – Peindre une image coup par coup
9	LSTM, GAN	Générateur sans input humain - Générer des paysages naturels évoquant des émotions
10	GAN	Générer des œuvres par requête (multimodal)
11	GAN, CNN, RNN	Générer des œuvres par requête (multimodal)

5. Expérimentation des algorithmes actuelles

5.1. Synthèse des protocoles d'expérimentation

Parmi les 11 articles sélectionnés, plusieurs protocoles d'expérimentation sont utilisés afin de déterminer la qualité du modèle ou algorithme proposé, ou afin de la comparer à d'autres modèles qu'on appellera Baseline.

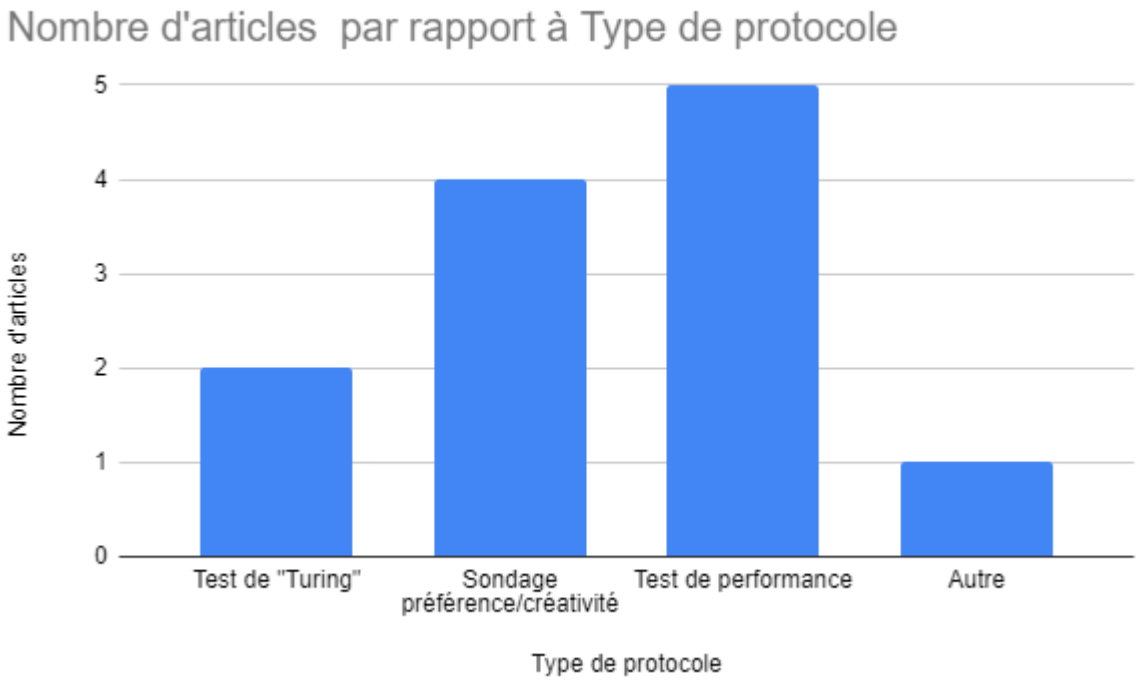

Figure 17 - Graphique du nombre d'articles par rapport aux types de tests

Il se distingue alors quatre types de protocoles d'expérimentation comme indiqué dans la figure 17 :

- **Un test de « Turing »** : originellement décrite par Alan Turing en 1950, ce test possède trois acteurs montrés sur la figure 18 : A, une machine, B un humain, et C un humain qui fera l'objet du test. A et B vont tous deux discuter avec C et fournir des réponses, et le test consiste à savoir quand est-ce que C ne pourra plus distinguer les réponses de la machine à celui de l'homme. Parmi les articles il existe plusieurs types de tests suivant un principe similaire. Le test

de Turing dans [2,1] consiste à demander à un sujet humain de déterminer si l'image mis devant lui a été créé par un homme ou une machine. Dans [1], on propose un protocole plus ressemblant au test de Turing originel, en mélangeant des œuvres humaines et des œuvres générées par des IA, puis de laisser des sujets humains les différencier.

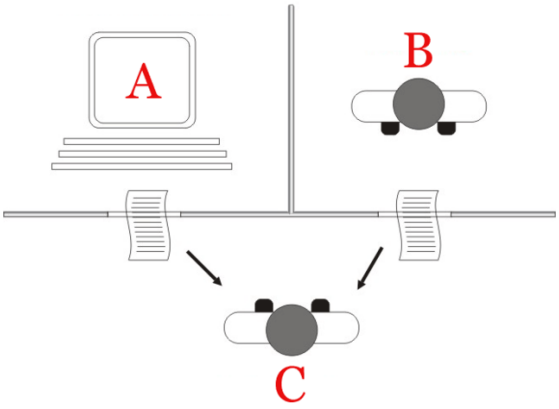

Figure 18 - Schéma du test de Turing : source [37]

- **Un sondage de préférence ou de critères artistiques** : 4 articles sur le set de 11 utilisent un sondage comme protocole d'expérimentation, dont le but est de demander à des individus leur jugement sur une œuvre qui lui est présentée. Il existe différents types de sondages ; dans [3,5] il y'a des sondages de « préférence » où l'on demande à des individus présentés avec deux images générées laquelle ils préfèrent, ou celle qu'ils trouvent la plus cohérente [3]. Ces tests de préférences sont effectués à la fois entre images générés par différents générateurs [3,5, 7], ou avec un générateur et des artistes humain, qui servent alors de Baseline [5,11]. Dans [5,6] le sondage est composé de critères : les individus donnent alors une note pour chaque critère qui seront recueilli en une moyenne pour être comparé à la baseline. Des exemples de critères sont : composition, créativité, clarté [5], compréhensible [3], cohérence de l'image selon une requête texte [7].

- **Un test de performance** : Le protocole expérimental le plus utilisé [figure 17] dans le set d'articles choisi est celui de ou des tests de performance, qui se déclinent selon des critères. Dans [6,7,8] le critère est celui de l'efficience où l'on mesure le modèle contre une Baseline d'autres algorithmes sur un même environnement et matériel. Dans [10] on utilise le test de la « Parzen Window » où l'on compare les résultats générés avec son data set : plus les différences sont grandes, plus cela signifie que l'algorithme a appris et ne se contente pas de faire des copies du data set. Dans [1] c'est un test de « Few Shots Classification », en effet, la classification étant importante dans un générateur d'art, cela consiste à comparer plusieurs algorithmes classifiant des data sets nouveaux, qui ne faisaient pas partis de l'entraînement. Ainsi, cela permet de juger la capacité réelle de classification d'un générateur, pour s'assurer qu'il ne soit pas trop dépendant de son data set d'entraînement.

- **D'autres types de protocoles d'expérimentation, plus rares** : Le modèle ERU proposé dans [9] vise à générer de l'émotion. Le protocole expérimental proposé est similaire à la catégorie sondage/préférence, mais sera mis à part car il ne vise pas à juger une œuvre, que ça soit d'une machine ou d'un artiste humain. L'objectif est de mesurer l'efficacité de l'ERU et de sa capacité de classification. Le protocole consiste à présenter une photo de paysage naturel, puis de demander à des utilisateurs de juger, sur une note sur 5, deux critères : la valence, qui représente le plaisir de l'utilisateur à visionner l'image, et « l'arousal » qui représente l'excitement. Plus l'arousal est bas, plus l'image serait calmante, ou sereine. Après avoir recueilli les notes utilisateurs, ils sont comparés aux valeurs données par le générateur lors de classification, et un intervalle d'erreur est donnée puis comparée à la Baseline.

  Le tableau 2 ci-dessous permet de synthétiser les protocoles expérimentaux utilisés dans le set d'articles. A noter qu'un article peut inclure plusieurs protocoles.

Tableau 2 - Synthèse des protocoles expérimentaux

Source	Echantillon	Protocole expérimentaux
1	300 images d'entraînement / 45 de test	Performance – Few shots classification
3	100 images à noter, 3 individus	Préférence/Sondage – Sondage sur 3 critères noté sur 4 : Understandable, Relevance, Veracity, et 3 critères noté de 1 ou 0 : Content, Form et context existence
5	18 images à noter par 242 participants, dont 29 parlant le mandarin nativement	Préférence/Sondage + Turing – 3 questions : 1/ L'image a-t-elle été créé par une machine ? 2/ Etes-vous sûr (Note de 1 à 10) 3/ Noter de 1 à 4 sur ces critères : aesthetically pleasing, artfully-composed, clear, creative
6	Semantic Boundaries Dataset en entraînement, GrabCut/DAVIS/Microsoft COCO en test	Performance – Test d'efficience (en combien d'étape l'image est segmentée)
7	Dataset : Microsoft COCO	Performance + Sondage – Test de performance contre StackGAN + Sondage

		sur l'image correspond le plus à une description/ la plus cohérente
8	100 images de WikiArt, 100 portraits de FFHQ, et 100 scènes naturelles de [38]	Performance – Test de d'efficience et précision contre des baselines avec une carte graphique NVIDIA 2080 Ti
9	Plateforme Amazon Mturk – Notation de 400 images par 1000 utilisateurs	Autre – Comparaison de la classification de l'algorithme et par des utilisateurs sur 2 critères : valence et arousal
10	Data set : CIFAR-10	Performance – Parzen Window
11	Plusieurs expérimentations ; échantillon de 18 à 100	Turing + Préférence – Question « Est-ce que cette image a été générée par une machine ou non » + Préférence entre des images de baseline ou entre machine / artiste humain

5.2. Data sets

Le processus le plus important d'un générateur d'art est l'entraînement des réseaux de neurones. Cet entraînement nécessite un data set d'image sur lequel apprendre, dont il existe plusieurs types comme montré dans la figure 19.

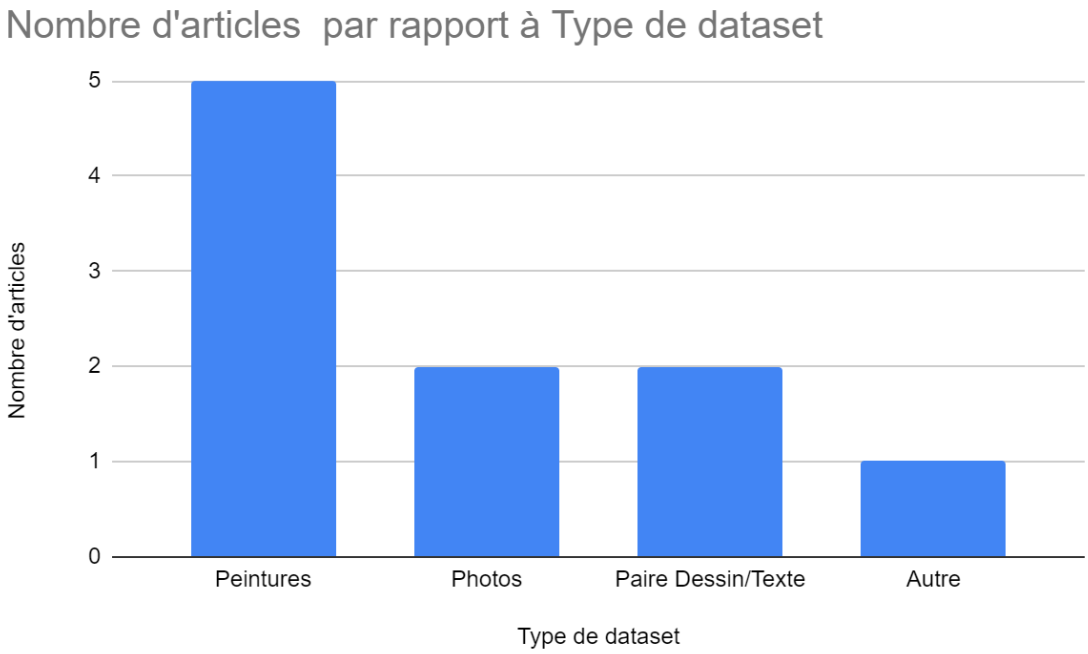

Figure 19 - Nombre d'articles selon le type de dataset

Les différents types de datas et utilisés dans les articles sont :

- Les peintures ou dessins de sources humaines sont le type de data set le plus utilisé parmi les articles, ce qui est peu surprenant pour des générateurs d'art. Parmi ces data sets, WikiArt est la source la plus utilisée et hormis [5] dont le but est de générer un style précis, les data sets préférés sont ceux contenant des œuvres de divers artistes ainsi que d'époque différente. Ce type de data set est utilisé dans les générateurs non-multimodaux.
- De façon similaire, si le générateur d'art cherche plutôt à générer des œuvres plutôt réalistes, ou inspirée par la photographie, il existe également des data sets contenant uniquement des photos, comme utilisé dans [7,9]
- Dans le cas de générateurs multimodaux, les data sets utilisés sont des paires d'image et de texte. Un exemple serait SemArt utilisé dans [3], ce data set ne

contient pas seulement des peintures, mais lient chacune d'entre elles avec une description, comme celle d'un musée, ainsi que de sept attributs (Ex : Artiste, Type, Période, Date...). L'utilisation d'une association image-texte est plus adaptée pour les générateurs multimodaux, qui s'entraînent à la fois sur la génération d'image, et la compréhension d'une requête texte.

- Le data set dans [1] utilise également des paires, mais cette fois d'image et d'esquisses. Ce type de data set est spécifique à l'objectif du générateur de [1], qui est d'apprendre à un générateur à reproduire une photo via des symboles.

Il existe donc plusieurs types de data sets selon les objectifs et besoins du générateur, et bien que certains modèles utilisent des data sets en open source, un data set spécifique peut être crée et utilisé dans le cas d'un modèle précis.

En outre du type, il existe des sous-groupes dans un même data set. En effet, il existe trois groupes au sein d'un data set :

- Le groupe « entraînement », généralement le groupe contenant le plus d'éléments, qui est utilisé afin d'entraîner un générateur. Par exemple, le data set ImageNet qui est utilisé dans [1], continent jusqu'à 1 millions d'images pour l'entraînement.
- Le groupe « validation », ce groupe est souvent celui contenant le moins d'éléments, ou selon le modèle, n'est pas toujours utilisé. Utilisé après l'entraînement, il permet d'ajuster les paramètres d'un générateur. Dans ImageNet, ce groupe est composé de 50 000 images.
- Le groupe de « test » correspond aux éléments utilisés pour tester un générateur. Il est différent du groupe d'entraînement afin que les générateurs ne soient pas biaisés, étant des éléments qui n'ont pas été vu auparavant. Dans [6], le protocole expérimental inclut plusieurs data sets de test afin de mesurer les performances du générateur sur la classification des images. Dans ImageNet, la taille du groupe de test est de 100 000 images.

Le tableau 3 ci-dessous permet de synthétiser les data sets utilisés dans le set d'articles. A noter qu'un article peut inclure plusieurs data sets.

Tableau 3 - Synthèse dataset

Source	Taille du Data set	Data set + Type
1	75k	ImageNet, images contenant des photos et leurs esquisses
3	21k	SemArt : Chaque image est commentée et possède 7 attributs. Inclut 1k d'images de tests et 1k pour la validation
4	179k	Trois sources (Web Gallery Art, WikiArt, BBC your painting), plusieurs époques et artistes
5	192	Peintures traditionnelles chinoises de plusieurs musées américains, rognées pour enlever les calligraphies
6	11k	Semantic Boundaries Dataset, 80% Training, 20% Test Autres data sets de test : GrabCut, DAVIS, Microsoft COCO
7	154k	Visual Genome (108k, Photos de paysages), COCO-Stuff (45k, photos de trains)
8	80k	WikiArt Peintures, 1119 artistes différents
9	10k	CGnA10766, images issues de photos diverses
10	80k	WikiArt, 80k, Peintures, 1119 artistes différents, 70% Training, 30% Test + CIFAR-

		10, 60K images 32x32px, 90% Training et 10% Test
11	80k | 10k	Data set 1 : WikiArt, Peintures, 1119 artistes différents Data set 2 : peintures, 50 artistes/mouvements différents

5.3. Synthèse-Résultats

Afin de conclure cette cinquième partie, ce chapitre aura pour but de synthétiser les résultats du set de 11 articles, selon les tests utilisés :

- Test de « Turing » : Parmi les cinq articles utilisant comme protocole un test de Turing, ou un procédé similaire, les résultats sont tous positifs. Dans le cas de [11], les individus discernent difficilement une œuvre générée d'une œuvre humaine, ce constat se retrouve dans [5]. Dans les résultats de [11], même en termes de préférence entre des œuvres générées ou humains, lorsqu'un individu n'en connaît pas les origines, le taux de préférence constate qu'il n'y a pas forcément une préférence pour l'un ou l'autre. Cependant, un des critères utilisés dans [5] montre qu'un défaut des œuvres générées par une IA est qu'ils ne sont pas forcément cohérents, ou qu'ils peuvent manque de clarté. Lorsque ces critères ne sont pas respectés, il devient alors aisé de différencier l'œuvre d'une machine d'une œuvre humaine ; si le style de l'œuvre n'est pas abstrait.

- Test de « préférence » : Les résultats d'un premier générateur dans [11] constatent une légère préférence pour les œuvres générées, par rapport aux œuvres humaines. Dans le cas du second générateur dans [11], aucune préférence particulière n'a été constatée, et en reliant avec les résultats des tests de Turing, il semblerait logique qu'il n'y ait pas de préférence particulière si l'individu ne sait pas différencier l'origine de l'œuvre. Cependant, une critique émise est que les individus interrogés pour le protocole expérimental ne font pas partis d'un milieu artistique ou ne sont pas particulièrement renseignés sur l'art. Dans [5], le protocole a inclus des personnes parlant mandarin couramment afin de déterminer si leurs avis différaient du reste de l'échantillon, mais la proximité à une culture reste un critère différent à celui d'un expert en art. Par ailleurs, les résultats en termes de préférence de [5] souligne un meilleur score pour les œuvres d'origine humaines, cela peut être dû au modèle ayant moins de data set que ceux dans [11], ou de la nature plus spécifique du générateur. Dans le cas de [3,7] où les résultats sont comparés à une baseline,

selon des critères à la correspondance d'une image générée à une requête humaine, les IA génératives permettent d'obtenir des résultats de plus en plus satisfaisants. Dans [7, 39] on peut souligner que malgré des manques de performance en termes de plaisance esthétique sur les générateurs actuels, ou des difficultés à générer des œuvres cohérentes, il y'a une amélioration des générateurs ; notamment grâce à la multimodalité et la supervision que celle-ci apporte.

- Test de performance : Les tests de performance sont divers mais on peut faire des liens en le type de test de performance, ainsi que le type de réseau de neurones artificiels utilisés. Les générateurs RNN, dû à leur nature récurrente, utilise plutôt des tests de performance, comme dans [1] ; les protocoles expérimentaux utilisés pour les GAN sont variés (performance, préférence, Turing) dû au fait que ça soit le type de réseau de neurones artificiels le plus utilisé dans la génération d'art du set d'articles. La mention du « Few Shots Classification » dans [1] souligne cependant des contraintes de performance lors de l'entraînement d'un générateur. Dans [10] la méthode du « Parzen Window » est utilisée afin de mesure la différence des images générées. En effet, si l'entraînement du générateur est mal effectué ou ne possède pas un data set suffisant, alors toutes les images qu'elle peut produire vont finir par trop se ressembler : c'est ce que l'on appelle « overfiting ».

Lorsqu'il s'agit de génération d'image, le GAN est le modèle le plus utilisé et semble être le plus performant [39] ; tandis que les autres types de réseaux de neurones permettent de constituer le GAN plutôt que d'être utilisé comme un générateur à part entière. Un défi du GAN est la quantité nécessaire de data set pour l'entraînement, qui peut être très grande, constituant indirectement un besoin pour des tests de performance sur ce sujet. Cette dépendance au data set pour l'entraînement est également un défi, car surentraîné le générateur ne pourra pas produire des images différentes du data set. Parmi le set d'articles, il semble que les protocoles de type « Test de Turing » obtiennent de bons résultats : il est plus

en plus difficile, pour un individu non-expert en art, de déterminer si une œuvre est d'origine humaine ou non. Cependant, dans [5] il existe encore des critères où les œuvres générées scorent plus bas que des œuvres d'humaines, notamment plus dans la notion de composition et d'esthétique, mais cette préférence n'est pas forcément corrélée à l'identification d'une œuvre comme étant générée par une machine.

Le tableau 4 ci-dessous permet de synthétiser les résultats du set d'articles.

Tableau 4 - Synthèse des résultats

Source	Protocole expérimentaux	Type de réseaux + Résultat
1	Performance – Few shots classification	RNN /CNN – Performe mieux que la baseline. La version RNN performe de façon significativement mieux que la version CNN.
3	Préférence/Sondage – Sondage sur 3 critères noté sur 4 : Understandable, Relevance, Veracity, et 3 critères noté de 1 ou 0 : Content, Form et context existence	GAN, LSTM – Performe mieux que la baseline hormis avec comme source de données Wikipedia, étant plus complexe. Conclut que pour classifier une image, il faut bien analyser l'image et les textes liés à l'artiste. L'entraînement est difficile dû au besoin d'un grand data set.
5	Préférence/Sondage + Turing – 3 questions : 1/ L'image a-t-elle été créé par une machine ?	GAN – Performe mieux que la Baseline dans les deux évaluations. Mais toutes les œuvres générées ont un

	2/ Etes-vous sûr (Note de 1 à 10) 3/ Noter de 1 à 4 sur ces critères : aesthetically pleasing, artfully-composed, clear, creative	score plus bas que les œuvres humaines sur la question 3.
6	Performance – Test d'efficience (en combien d'étape l'image est segmentée)	GAN – Performe mieux que la Baseline à la fois en efficience et rapidité.
7	Performance + Sondage – Test de performance contre StackGAN + Sondage sur l'image correspond le plus à une description/ la plus cohérente	GAN – Performe mieux que StackGAN sur tous les critères grâce aux requêtes en graphe (scene graph) VS du texte non structuré
8	Performance – Test de d'efficience et précision contre des Baseline avec une carte graphique NVIDIA 2080 Ti	CNN – En termes d'efficience, performe légèrement mieux que la baseline, mais requiert beaucoup moins d'entraînement
9	Autre – Comparaison de la classification de l'algorithme et par des utilisateurs sur 2 critères : valence et arousal	LSTM, GAN – Performe mieux que la baseline mais l'objet d'étude (paysage naturel) est limité, et le générateur perd en

		cohérence en se concentrant sur la création d'émotion
10	Performance – Parzen Window	GAN – Performe mieux que baseline mais il existe des difficultés concernant l'évolution d'un artiste du dataset (va ignorer l'évolution du style), ou des styles plus spécifiques (ukiyo-e, qui est originellement peint sur du bois)
11	Turing + Préférence – Question « Est-ce que cette image a été générée par une machine ou non » + Préférence entre des images de baseline ou entre machine / artiste humain	GAN – 1. 63.3% des participants préfèrent l'image générée. 2. 53% des participants pensent que les œuvres générées sont d'origine humaines. Pas de différence de préférence entre les œuvres.

6. Conclusion

En conclusion, cet état de l'art montre qu'il y'a une avancée concernant l'utilisation de réseaux de neurones artificiels pour générer de l'art ; et plus particulièrement de l'art visuel. Ces générateurs se sont popularisés depuis la découverte de OpenAI par le public, et continue d'être développés à la fois dans le domaine artistique mais aussi dans d'autres domaines (exemple : ChatGPT).

Cette avancée est liée à la progression des modèles de neurones artificiels, notamment le GAN, et à l'amélioration de la capacité calculatoire des machines afin d'entraîner les générateurs sur des data sets très larges. Les générateurs connus du grand public sont ceux comportant également du NLP, afin de générer des œuvres via une requête texte. Cependant, comme noté dans [11], s'il est plus difficile de distinguer la nature humaine ou non d'une œuvre, les œuvres générées ne sont pas particulièrement plus appréciées que les œuvres humaines ou même moins appréciées comme dans [5].

Les futurs défis de la génération d'art visuel sembleraient être donc :

- La quantité de data set nécessaire pour l'entraînement. Cette quantité engendre naturellement des contraintes de performance sur le hardware mais aussi dans la qualité de l'algorithme.
- La qualité esthétique de l'art généré. Ce critère étant difficile à quantifier, il serait possible de faire appel à des vis d'experts lors des expérimentations, comme souligné dans [11]. Ce point m'intéresse pour le mémoire de M2 afin de travailler sur un protocole expérimental qui prendrait en compte des avis de personnes expertes, ou du moins sensibles à l'art, en interrogeant des personnes amateures d'arts, ou étudiant/travaillant dans ce domaine. Cela permettrait de souligner des critères plus précises la qualité de l'art générative.
- L'éthique du générateur. En effet, les générateurs sont entraînés sur des œuvres humaines dont les auteurs peuvent ne pas avoir consentis à l'utilisation de leurs images. De ce fait, cela soulève une question de droit d'auteur, ou d'éthique puisque certains générateurs peuvent être utilisés à de fin lucratifs en s'étant entraîné sur des artistes n'y ayant pas consentis.

Table des figures

Table des tableaux

Bibliographie

1. Wang, Alexander, Mengye Ren, and Richard Zemel. "Sketchembednet: Learning novel concepts by imitating drawings." In *International Conference on Machine Learning*, pp. 10870-10881. PMLR, 2021
2. Colton, Simon, Amy Smith, Sebastian Berns, Ryan Murdock, and Michael Cook. "Generative search engines: Initial experiments." In *Proceedings of the International Conference on Computational Creativity*. 2021 == Source non-fiable mais potentiel pour l'introduction (les différentes utilisations d'art généré par des IA)
3. Bai, Zechen, Yuta Nakashima, and Noa Garcia. "Explain me the painting: Multi-topic knowledgeable art description generation." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5422-5432. 2021.
4. Lee, Byunghwee, Daniel Kim, Seunghye Sun, Hawoong Jeong, and Juyong Park. "Heterogeneity in chromatic distance in images and characterization of massive painting data set." *PLoS One* 13, no. 9 (2018): e0204430.
5. Xue, Alice. "End-to-end Chinese landscape painting creation using generative adversarial networks." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3863-3871. 2021.

6. Li, Zhuwen, Qifeng Chen, and Vladlen Koltun. "Interactive image segmentation with latent diversity." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 577-585. 2018.
7. Johnson, Justin, Agrim Gupta, and Li Fei-Fei. "Image generation from scene graphs." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1219-1228. 2018.
8. Liu, Songhua, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. "Paint transformer: Feed forward neural painting with stroke prediction." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6598-6607. 2021.
9. Park, C. and Lee, I.K., 2020. Emotional Landscape Image Generation Using Generative Adversarial Networks. In *Proceedings of the Asian Conference on Computer Vision*.
10. Tan, Wei Ren, Chee Seng Chan, Hernán E. Aguirre, and Kiyoshi Tanaka. "ArtGAN: Artwork synthesis with conditional categorical GANs." In 2017 IEEE International Conference on Image Processing (ICIP), pp. 3760-3764. IEEE, 2017.
11. Shahriar, Sakib. "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." *Displays* (2022): 102237.
12. A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, “Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms,” *arXiv preprint arXiv:1706.07068*, 2017.
13. J Shane. Sea shanty surrealism, 2021. aiweirdness.com/post/645282704595795968/ sea-shanty-surrealism.
14. A Smith and S Colton. CLIP-Guided GAN Image Generation: An Artistic Exploration. In *Proceedings of the EvoMusArt conference*, 2021.
15. China bans AI-generated media without watermarks | *Ars Technica*, 2022
16. 国家互联网信息办公室等三部门发布《互联网信息服务深度合成管理规定》-中共中央网络安全和信息化委员会办公室 (cac.gov.cn), 2022

17. Ivakhnenko, Alekseĭ Grigor′evich, Valentin Grigorevich Lapa, and Valentin Grigor′evich Lapa. *Cybernetics and forecasting techniques. Vol.* 8. American Elsevier Publishing Company, 1967.
18. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60, no. 6 (2017): 84-90.
19. Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in Neural Information Processing Systems 33* (2020): 6840-6851.
20. "GAN 2.0: NVIDIA's Hyperrealistic Face Generator". *SyncedReview.com*. December 14, 2018. Retrieved October 3, 2019.
21. Roose, Kevin (September 2, 2022). "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy". The New York Times. Retrieved September 2, 2022.
22. Vincent, James (January 16, 2023). "AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit". *The Verge.*
23. N. Xu, B. L. Price, S. Cohen, J. Yang, and T. S. Huang. *Deep interactive object selection*. In CVPR, 2016.
24. Xie, Saining, and Zhuowen Tu. "Holistically-nested edge detection." In *Proceedings of the IEEE international conference on computer vision,* pp. 1395-1403. 2015.
25. Zhi, Jiale. "Pixelbrush: Art generation from text with gans." Cl. *Proj. Stanford CS231N Convolutional Neural Networks Vis. Recognition*, Sprint 256 (2017).
26. Ruan, Shulan, Yong Zhang, Kun Zhang, Yanbo Fan, Fan Tang, Qi Liu, and Enhong Chen. "Dae-gan: Dynamic aspect-aware gan for text-to-image synthesis." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13960-13969. 2021.
27. Gogul, I., and V. Sathiesh Kumar. "Flower species recognition system using convolution neural networks and transfer learning." In *2017 fourth international conference on signal processing, communication and networking (ICSCN)*, pp. 1-6. IEEE, 2017.

28. Sak, Haşim, Andrew Senior, and Françoise Beaufays. "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition." *arXiv preprint arXiv*:1402.1128 (2014).
29. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation 9, no. 8* (1997): 1735-1780.
30. *https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks*
31. Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *In Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1-10*).*
32. https://commons.wikimedia.org/wiki/File:Neural_network.svg?uselang=fr
33. https://developers.google.com/machine-learning/gan/gan_structure?hl=fr
34. Petersen, Kai, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. "Systematic mapping studies in software engineering." In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)* 12, pp. 1-10. 2008.
35. Srinivasan, Ramya, and Kanji Uchino. "Biases in generative art: A causal look from the lens of art history." *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 41-51. 2021.
36. Boden, Margaret A., and Ernest A. Edmonds. *"What is generative art?." Digital Creativity 20*, no. 1-2 (2009): 21-46.
37. https://en.wikipedia.org/wiki/Turing_test#/media/File:Turing_test_diagram.png
38. https://www.kaggle.com/datasets/arnaud58/landscape-pictures

39. Wang, Lei, Wei Chen, Wenjia Yang, Fangming Bi, and Fei Richard Yu. "A state-of-the-art review on image synthesis with generative adversarial networks." *IEEE Access 8* (2020): 63514-63537.

Annexes

1. Liste d'articles rejetés

Articles rejetés pour sources non-fiables :

- Quan, Huafeng, Shaobo Li, and Jianjun Hu. "Product innovation design based on deep learning and Kansei engineering." *Applied Sciences 8, no. 12* (2018): 2397.
- Liu, Vivian, Han Qiao, and Lydia Chilton. "Opal: Multimodal Image Generation for News Illustration." *arXiv preprint arXiv:2204.09007* (2022).
- Conwell, Colin, and Tomer Ullman. "Testing relational understanding in text-guided image generation." *arXiv preprint arXiv:2208.00005* (2022).
- Seneviratne, Sachith, Damith Senanayake, Sanka Rasnayaka, Rajith Vidanaarachchi, and Jason Thompson. "DALLE-URBAN: Capturing the urban design expertise of large text to image transformers." *arXiv preprint arXiv:2208.04139* (2022).
- CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms
- Im, Daniel Jiwoong, Chris Dongjoo Kim, Hui Jiang, and Roland Memisevic. "Generating images with recurrent adversarial networks." *arXiv preprint arXiv:1602.05110* (2016).
- Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:*1508.06576 (2015).

Articles rejetés pour manque de pertinence/hors sujet :

- Kong, Shu, and Deva Ramanan. "Opengan: Open-set recognition via open data generation." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 813-822. 2021.

- Zhou, Yufan, Ruiyi Zhang, Changyou Chen, Chunyuan Li, Chris Tensmeyer, Tong Yu, Jiuxiang Gu, Jinhui Xu, and Tong Sun. "Towards Language-Free Training for Text-to-Image Generation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17907-17917. 2022.
- Johnson, Rie, and Tong Zhang. "Composite functional gradient learning of generative adversarial models." In *International Conference on Machine Learning*, pp. 2371-2379. PMLR, 2018.
- Tewel, Yoad, Yoav Shalev, Idan Schwartz, and Lior Wolf. "ZeroCap: Zero-Shot Image-to-Text Generation for Visual-Semantic Arithmetic." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 17918-17928. 2022.
- Xu, Tao, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1316-1324. 2018

Articles sans accès :

- https://ieeexplore.ieee.org/abstract/document/9175557
- https://ieeexplore.ieee.org/abstract/document/9540115
- https://ieeexplore.ieee.org/abstract/document/9540081
- https://ieeexplore.ieee.org/abstract/document/9796760
- https://ieeexplore.ieee.org/abstract/document/8296985

2. Tableau de synthèse

N° Article	Article	Algorithmes/Modèles	Dataset	Expérimentation
1	Shahriar, Sakib. "GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network." Displays (2022): 102237	1 : GAN + Encodeur pour classifier les peintures, 2: DCGAN	Dataset 1 : WikiArt, 80k, Peintures, 1119 artistes différents, Dataset 2: 10k, peintures, 50 artistes/mouvements différents	1 : Préférence entre les images générées & la baseline n=100 63.3% des participants préfèrent l'image générée 2: Sondage : Test de Turing + préférence noté sur 5 n= 18 53% turing / 3.2 en préférence vs 3.1 pour les vrais artistes
2	Xue, Alice. "End-to-end Chinese landscape painting creation using generative adversarial networks." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3863-3871. 2021.	SketchGAN, PaintGAN	192 peintures traditionnelles de différents musées américains	Sondage : Test de Turing + Préférence. Turing consiste à demander la confiance de l'individu sur l'origine de l'oeuvre, préférence repose sur 4 critères : esthétique, clareté, composition, créativité
3	Liu, Songhua, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. "Paint transformer: Feed forward neural painting with stroke prediction." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 6598-6607. 2021	CNN, encodeur/décodeur, DETR	WikiArt, 80k, Peintures, 1119 artistes différents	Test de performance contre des baselines, sur une nvidia 2080Ti, Test de précision et efficience contre des baselines
4 (Article dédié pour l'introduction)	Colton, Simon, Amy Smith, Sebastian Berns, Ryan Murdock, and Michael Cook. "Generative search engines: Initial experiments." In Proceedings of the International Conference on Computational Creativity. 2021	NLP, GAN, encodeurs, CLIP model	//	//
5	Wang, Alexander, Mengye Ren, and Richard Zemel. "Sketchembednet: Learning novel concepts by imitating drawings." In International Conference on Machine Learning, pp. 10870-10881. PMLR, 2021	Encodeur CNN, Décodeur RNN, SketchEmbedded	ImageNet, 75471 images contenant des photos et des esquisses	Test de Few Shots Classification contre des baselines (classer des datasets inconnus lors du training)

6	Bai, Zechen, Yuta Nakashima, and Noa Garcia. "Explain me the painting: Multi-topic knowledgeable art description generation." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5422-5432. 2021.	Masked Sentence Generation, LSTM Decoder, NLP	SemArt dataset, 21k, Chaque image est commenté et possède 7 attributs. Inclut 1k d'images de tests et 1k pour la validation	Sondages sur 100 images et classifications, jugés sur plusieurs critères notés de 3 individus
7 (Article à utiliser avec les algos NLP et multimodale)	Lee, Byunghwee, Daniel Kim, Seunghye Sun, Hawoong Jeong, and Juyong Park. "Heterogeneity in chromatic distance in images and characterization of massive painting data set." PLoS One 13, no. 9 (2018): e0204430.	Metropolis-Hastings proces	179k Total de trois sources (Web Gallery Art, WikiArt, BBC your painting), plusieurs époques et artistes	//
8	Li, Zhuwen, Qifeng Chen, and Vladlen Koltun. "Interactive image segmentation with latent diversity." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 577-585. 2018.	GAN	Semantic Boundaries Dataset, 11k, 80% Training, 20% Test, Autres datasets de test : GrabCut, DAVIS, Microsoft COCO	Test de performance : Comparaison avec des 7 baselines sur l'efficience de la segmentation d'une image
9	Johnson, Justin, Agrim Gupta, and Li Fei-Fei. "Image generation from scene graphs." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1219-1228. 2018.	GAN (StyleGAN2), GNN	Visual Genome (108k, Photos de paysages), COCO-Stuff (45k, photos de train)	Test de performance avec comme baseline VisualGAN, Sondage avec des individus choisissant l'image la plus cohérente selon une description donnée, avec comme baseline StackGAN
10	Park, C. and Lee, I.K., 2020. Emotional Landscape Image Generation Using Generative Adversarial Networks. In Proceedings of the Asian Conference on Computer Vision.	LSTM,GAN	CGnA10766, 10k d'images issus de photos diverses	Comparaison des valeurs d'émotions donnée par l'algorithme VS ceux de l'échantillon humain (Amazon Mturk platform, n=1k pour 400 images)
11	Tan, Wei Ren, Chee Seng Chan, Hernán E. Aguirre, and Kiyoshi Tanaka. "ArtGAN: Artwork synthesis with conditional categorical GANs." In 2017 IEEE International Conference on Image Processing (ICIP), pp. 3760-3764. IEEE, 2017.	artGAN	WikiArt, 80k, Peintures, 1119 artistes différents, 70% Training, 30% Test + CIFAR-10, 60K images 32x32px, 90% Training et 10% Test	Test Parzen Window (Différences entre les résultats + datasets) contre baselines (DCGAN, GAE/VAE)