



HAL
open science

Les Generative Adversarial Networks

Moez Krichen

► **To cite this version:**

| Moez Krichen. Les Generative Adversarial Networks. 2022. hal-04140057

HAL Id: hal-04140057

<https://hal.science/hal-04140057v1>

Preprint submitted on 24 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les Generative Adversarial Networks

Moez Krichen

Laboratoire ReDCAD, Université de Sfax, Tunisie
moez.krichen@redcad.org

Résumé. Les Generative Adversarial Networks (GANs) sont des techniques d'apprentissage profond qui ont connu un succès remarquable dans la génération d'images, de vidéos et d'autres types de données réalistes. Cet article fournit un guide complet sur les GANs, couvrant leur architecture, leurs fonctions de perte, leurs méthodes d'entraînement, leurs applications, leurs métriques d'évaluation, leurs défis et leurs orientations futures. Nous commençons par une introduction aux GANs et à leur développement historique, suivie d'une revue de la littérature et des travaux connexes. Nous fournissons ensuite un aperçu détaillé de l'architecture des GANs, y compris les réseaux générateurs et discriminateurs, et discutons des choix de conception clés et des variations. Ensuite, nous examinons les fonctions de perte utilisées dans les GANs, y compris l'objectif minimax original, ainsi que des approches plus récentes telles que la distance de Wasserstein et la pénalité de gradient. Nous plongeons ensuite dans l'entraînement des GANs, en discutant des techniques courantes telles que l'optimisation alternée, la discrimination par mini-batch et la normalisation spectrale. Nous fournissons également une enquête sur les différentes applications des GANs dans divers domaines. De plus, nous examinons les métriques d'évaluation utilisées pour évaluer la diversité et la qualité des données produites par les GANs. En outre, nous discutons des défis et des problèmes ouverts dans les GANs, y compris l'effondrement de mode, l'instabilité de l'entraînement et les considérations éthiques. Enfin, nous donnons un aperçu des orientations futures de la recherche sur les GANs, notamment l'amélioration de la scalabilité, le développement de nouvelles architectures, l'incorporation de connaissances de domaine et l'exploration de nouvelles applications. Dans l'ensemble, cet article sert de guide complet sur les GANs, fournissant des informations théoriques et pratiques pour les chercheurs et les praticiens dans le domaine.

1 Introduction

Le développement rapide des disciplines de l'apprentissage automatique (ML) et de l'apprentissage profond (DL) a conduit à des avancées significatives dans l'intelligence artificielle (AI) (16). L'objectif de ML, une branche de l'intelligence artificielle, est de créer des algorithmes qui peuvent apprendre automatiquement à partir de données et générer des prédictions ou des jugements.

En revanche, DL est un sous-ensemble de ML qui utilise des réseaux neuronaux (NN) avec plusieurs couches pour apprendre des représentations de données complexes (2). En tant que forme de DL, les GAN ont démontré un succès remarquable dans la production de données de haute qualité et sont devenus un instrument crucial pour la synthèse et l'amélioration des données.

En raison de sa capacité à apprendre des représentations de données complexes, DL a révolutionné diverses disciplines, notamment la vision par ordinateur, le traitement du langage naturel (NLP)¹, et la reconnaissance vocale. Les modèles DL, y compris les CNN² et les RNN³, ont atteint des performances de pointe sur diverses tâches difficiles, telles que la reconnaissance d'images et la traduction de langues. Les GAN, qui utilisent une forme spécifique d'architecture de réseau neuronal, ont créé de nouvelles opportunités pour la génération et la synthèse de données et ont été utilisés pour produire des images, des vidéos et de l'audio réalistes.

Dans ce document, nous nous concentrons spécifiquement sur les réseaux antagonistes génératifs (GAN), qui sont une forme de DL, et fournissons un guide complet de leur architecture, des fonctions de perte, des méthodes d'entraînement, des applications, des métriques d'évaluation, des défis et des orientations futures. Nous croyons que les GAN représentent une avancée significative dans le domaine de l'AI et ont le potentiel de débloquent de nouvelles opportunités de découverte scientifique et d'expression artistique. En fournissant un aperçu complet des GAN, nous espérons rendre ce sujet complexe et difficile plus accessible aux chercheurs et praticiens du domaine et inspirer de nouvelles innovations et applications.

Les GAN ont émergé comme une technique utile pour produire des données réalistes dans diverses disciplines allant de la vision par ordinateur et des graphiques à NLP et la synthèse audio. Goodfellow et al. ont développé les GAN en 2014, et depuis lors, ils ont été l'un des domaines de recherche les plus actifs en DL. Les GAN sont composés de deux réseaux neuronaux adversaires, un réseau générateur (*Generator_{Network}*) et un réseau discriminatoire (*Discriminator_{Network}*). Le *Generator_{Network}* apprend à produire des données indiscernables des données réelles (*Real_{Data}*), tandis que le *Discriminator_{Network}* apprend à différencier entre les deux.

Malgré leur succès, les GAN sont un sujet complexe et difficile qui nécessite une compréhension approfondie à la fois de la théorie sous-jacente et de la mise en œuvre pratique. Dans ce document, nous fournissons un guide complet des GAN, couvrant leur architecture, les fonctions de perte, les méthodes d'entraînement, les applications, les métriques d'évaluation, les défis et les orientations futures. Notre objectif est de fournir des connaissances théoriques et pratiques aux chercheurs et praticiens du domaine et de contribuer à démystifier les aspects souvent confus et intimidants des GAN.

Les principales contributions de ce document sont les suivantes :

1. Une vue d'ensemble complète de l'architecture des GAN, y compris le *Generator_{Network}* et le *Discriminator_{Network}*, et les choix de conception clés et les variations.

1. NLP = Natural Language Processing
2. CNN = Convolutional Neural Network
3. RNN = Recurrent Neural Network

2. Une revue approfondie des fonctions de perte utilisées dans les GAN, y compris l'objectif minimax original, ainsi que des approches plus récentes telles que la distance de Wasserstein et la pénalité de gradient.
3. Une enquête sur les différentes méthodes d'entraînement utilisées dans les GAN, y compris l'optimisation alternée, la discrimination des mini-batches et la normalisation spectrale.
4. Un examen des différentes applications des GAN dans des domaines tels que la vision par ordinateur, NLP et la synthèse audio.
5. Une exploration des métriques d'évaluation utilisées pour évaluer la diversité et la qualité des données produites par les GAN, telles que la distance de Fréchet Inception et le score Inception.
6. Une discussion des défis et des problèmes ouverts dans la recherche sur les GAN, notamment l'instabilité de l'entraînement, l'effondrement du mode et les considérations éthiques.
7. Un aperçu des orientations futures de la recherche sur les GAN, notamment l'amélioration de la scalabilité, le développement de nouvelles architectures, l'intégration de connaissances dans le domaine et l'exploration de nouvelles applications.

Dans la section 2, nous fournissons un bref historique sur les GAN et les travaux connexes. Dans la section 3, nous présentons un aperçu détaillé de l'architecture des GAN, y compris le *Generator_{Network}* et le *Discriminator_{Network}*, ainsi que les choix de conception clés et les variations. Dans la section 4, nous examinons les fonctions de perte utilisées dans les GAN, y compris l'objectif minimax original, ainsi que des approches plus récentes telles que la distance de Wasserstein et la pénalité de gradient. Dans la section 5, nous discutons des méthodes d'entraînement utilisées dans les GAN, y compris l'optimisation alternée, la discrimination des mini-batches et la normalisation spectrale. Dans la section 6, nous enquêtons sur les différentes applications des GAN dans des domaines tels que la vision par ordinateur, NLP et la synthèse audio. Dans la section 7, nous explorons les métriques d'évaluation utilisées pour évaluer la diversité et la qualité des données produites par les GAN, telles que la distance de Fréchet Inception et le score Inception. Dans la section 8, nous discutons des défis et des problèmes ouverts dans la recherche sur les GAN, notamment l'instabilité de l'entraînement, l'effondrement du mode et les considérations éthiques. Enfin, dans la section 9, nous donnons un aperçu des orientations futures de la recherche sur les GAN, notamment l'amélioration de la scalabilité, le développement de nouvelles architectures, l'intégration de connaissances dans le domaine et l'exploration de nouvelles applications. Nous concluons le document dans la section 10, en résumant nos contributions et en discutant de l'impact plus large et du potentiel des GAN.

2 Contexte

Ces dernières années, l'apprentissage profond (DL) est devenu un outil important pour résoudre une grande variété de problèmes d'apprentissage automatique, tels que la

classification d'images, la reconnaissance de la parole et le traitement du langage naturel (22; 7). Les algorithmes de DL reposent sur des réseaux de neurones (NN), qui sont composés de couches de nœuds de traitement interconnectés qui peuvent apprendre à reconnaître des motifs dans les données grâce à un processus d'entraînement supervisé ou non supervisé.

Les modèles génératifs sont une classe d'algorithmes de DL qui peuvent produire de nouvelles données similaires aux données d'entraînement (*TrainingData*). Ils ont de nombreux types d'applications, de la synthèse d'images à la génération de discours. L'un des types de modèles génératifs les plus populaires est le GAN.

Le concept de base des GAN a été introduit par Ian Goodfellow et ses collègues en 2014 (8). Les GAN sont composés de deux réseaux de neurones : un *GeneratorNetwork* et un *DiscriminatorNetwork*. Le *GeneratorNetwork* prend en entrée un vecteur de bruit aléatoire et produit un nouvel échantillon qui est censé être similaire aux données d'entraînement. Le *DiscriminatorNetwork* prend en entrée un échantillon et essaie de différencier les échantillons produits par le *GeneratorNetwork* des échantillons provenant des données d'entraînement. Le *GeneratorNetwork* est entraîné à produire des échantillons difficiles à différencier des données d'entraînement par le *DiscriminatorNetwork*, tandis que le *DiscriminatorNetwork* est entraîné à classifier correctement les échantillons comme étant soit réels soit faux. Le processus d'entraînement (*TrainingProcess*) pour les GAN est itératif et implique une alternance entre l'entraînement du *GeneratorNetwork* et du *DiscriminatorNetwork*. Pendant l'entraînement, le *GeneratorNetwork* apprend à produire des échantillons plus réalistes, tandis que le *DiscriminatorNetwork* apprend à devenir plus précis dans la différenciation entre les échantillons réels et faux. L'objectif est de trouver un équilibre où le *GeneratorNetwork* produit des échantillons indiscernables des données d'entraînement, et le *DiscriminatorNetwork* n'est pas en mesure de différencier les échantillons réels des faux.

Plusieurs types d'architectures GAN ont été proposés, notamment les GAN convolutionnels profonds, les WGAN (1) et les GAN conditionnels (6). Les DCGAN sont un type de GAN qui utilisent des CNN dans le *GeneratorNetwork* et le *DiscriminatorNetwork* pour produire des images de haute qualité. Les WGAN sont un type de GAN qui utilisent la distance de Wasserstein au lieu de la divergence Jensen-Shannon traditionnelle pour évaluer la distance entre les distributions produites et réelles. Les cGAN sont un type de GAN qui conditionnent le *GeneratorNetwork* et le *DiscriminatorNetwork* sur des informations supplémentaires, telles que des étiquettes de classe ou des vecteurs d'attributs.

En plus de la synthèse d'images, les GAN ont été appliqués à une large gamme de problèmes, tels que l'augmentation de données, le transfert de style (5) et la détection d'anomalies (32). La synthèse d'images basée sur les GAN a connu des avancées importantes ces dernières années, avec l'introduction de GAN progressifs, styleGAN (9) et BigGAN. Ces modèles sont capables de créer des images de haute qualité avec une résolution élevée et des styles variés.

3 Architecture GAN

Les GAN (8) sont un type de modèle génératif qui apprend à générer de nouveaux échantillons de données qui ressemblent à un $Training_{Data}$ donné. L'architecture de base des GAN se compose de deux réseaux de neurones : un $Generator_{Network}$ et un $Discriminator_{Network}$. Le $Generator_{Network}$ prend en compte un vecteur de bruit aléatoire $v \in \mathbb{R}^d$ en entrée et crée un échantillon synthétique ($Synthetic_{Data}$) $\hat{s} \in \mathbb{R}^m$ en sortie. Le $Discriminator_{Network}$ prend un échantillon de données $s \in \mathbb{R}^m$ en entrée et produit une valeur scalaire $D(s) \in [0, 1]$ en sortie, indiquant la probabilité que s soit un échantillon de $Real_{Data}$ (par opposition à un échantillon synthétique produit par le $Generator_{Network}$).

Le $Generator_{Network}$ et le $Discriminator_{Network}$ sont entraînés de manière adversaire, le $Generator_{Network}$ tentant de générer des échantillons synthétiques qui ressemblent aux échantillons de $Real_{Data}$ et le $Discriminator_{Network}$ tentant de différencier les échantillons de $Real_{Data}$ et de $Synthetic_{Data}$. Le $Training_{Process}$ peut être modélisé comme un jeu minimax à deux joueurs dans lequel le $Generator_{Network}$ G tente de minimiser la fonction objectif suivante :

$$\min_G \max_D \mathbb{E}_{s \sim p_{data}(s)} [\log D(s)] + \mathbb{E}_{v \sim p_v(v)} [\log(1 - D(G(v)))], \quad (1)$$

où :

- $p_{data}(s)$ = distribution réelle des données ;
- v = vecteur de bruit ;
- $p_v(v)$ = distribution priori de v ;
- \mathbb{E} = la valeur attendue.

Le premier terme dans l'équation 1 encourage le $Discriminator_{Network}$ à classifier correctement les échantillons de $Real_{Data}$ comme étant réels, tandis que le second terme encourage le $Generator_{Network}$ à générer des échantillons synthétiques que le $Discriminator_{Network}$ classe comme réels.

Le $Discriminator_{Network}$ D tente de maximiser la fonction objectif suivante :

$$\max_D \mathbb{E}_{s \sim p_{data}(s)} [\log D(s)] + \mathbb{E}_{v \sim p_v(v)} [\log(1 - D(G(v)))], \quad (2)$$

où le premier terme encourage le $Discriminator_{Network}$ à classifier correctement les échantillons de $Real_{Data}$ comme étant réels, tandis que le second terme encourage le $Discriminator_{Network}$ à classifier correctement les échantillons synthétiques comme étant faux.

Dans l'ensemble, les GAN ont montré un succès remarquable dans la génération d'échantillons de $Synthetic_{Data}$ de haute qualité dans une variété de domaines, notamment les images, l'audio et le texte. Les différents types d'architectures GAN continuent d'évoluer et de s'améliorer, et ont un grand potentiel pour des avancées futures dans la modélisation générative.

4 Fonctions de perte pour les GAN

Le succès des GAN dans la génération d'échantillons de $Synthetic_{Data}$ de haute qualité est étroitement lié à la conception de leurs fonctions de perte ($Loss_{Function}$). Dans cette section, nous passons en revue certaines des fonctions de perte les plus couramment utilisées pour les GAN et leurs propriétés.

4.1 La fonction de perte originale des GAN

La fonction de perte originale des GAN (8) est donnée par l'équation 1, qui encourage le $Generator_{Network}$ G à générer des échantillons synthétiques indiscernables des échantillons réels par le $Discriminator_{Network}$ D . Bien que la fonction de perte originale des GAN ait été couronnée de succès dans la génération d'échantillons de $Synthetic_{Data}$ de haute qualité, elle souffre de plusieurs problèmes, notamment l'instabilité lors de l'entraînement et l'effondrement de mode, où le $Generator_{Network}$ apprend à produire un ensemble limité d'échantillons qui ne représentent pas toute la diversité de la distribution de données réelle.

4.2 Fonctions de perte améliorées des GAN

Pour résoudre les problèmes de la fonction de perte originale des GAN, plusieurs fonctions de perte améliorées des GAN ont été proposées dans la littérature.

Les GAN de Wasserstein (WGAN) (1) utilisent la distance de Wasserstein comme fonction de perte, qui a été démontrée pour produire un entraînement plus stable et des échantillons de meilleure qualité. La fonction de perte WGAN est donnée par l'équation 3 :

$$\min_G \max_D \mathbb{E}_{s \sim p_{data}(s)} [D(s)] - \mathbb{E}_{v \sim p_v(v)} [D(G(v))], \quad (3)$$

où D est une fonction 1-Lipschitz, et le $Discriminator_{Network}$ D est entraîné pour maximiser l'équation 3.

Une autre approche consiste à utiliser une fonction de perte des moindres carrés, comme proposé dans les GAN des moindres carrés (LSGAN) (28). La fonction de perte LSGAN est donnée par l'équation 4 :

$$\min_G \max_D \frac{1}{2} \mathbb{E}_{s \sim p_{data}(s)} [(D(s) - 1)^2] + \frac{1}{2} \mathbb{E}_{v \sim p_v(v)} [D(G(v))^2], \quad (4)$$

où le $Discriminator_{Network}$ D est entraîné pour minimiser l'équation 4.

D'autres approches comprennent la fonction de perte de charnière utilisée dans les GAN de charnière (HingeGANs) (33) et la fonction de perte de correspondance de caractéristiques utilisée dans les GAN de correspondance de caractéristiques (Feature Matching GANs) (30).

En conclusion, le choix de la fonction de perte est critique pour le succès des GAN dans la génération d'échantillons de $Synthetic_{Data}$ de haute qualité. Bien que la fonction de perte originale des GAN ait été couronnée de succès dans de nombreuses applications, plusieurs fonctions de perte améliorées ont été proposées pour résoudre ses

limitations et produire un entraînement plus stable et des échantillons de meilleure qualité. Le type de fonction de perte utilisé est déterminé par le but et le problème à résoudre.

5 Le Training des GANs

Les GAN sont généralement entraînés en utilisant un jeu à deux joueurs minimax, où un *Generator_{Network}* apprend à produire des échantillons de *Synthetic_{Data}*, et un *Discriminator_{Network}* apprend à différencier les échantillons de *Real_{Data}* des échantillons de *Synthetic_{Data}*. Le *Training_{Process}* implique de mettre à jour de manière itérative les paramètres du *Generator_{Network}* et du *Discriminator_{Network}* pour améliorer leurs performances.

5.1 Défis de l'entraînement des GAN

L'entraînement des GAN peut être difficile en raison de plusieurs facteurs, notamment l'instabilité et l'effondrement de mode. L'instabilité peut survenir lorsque le *Discriminator_{Network}* est trop puissant et apprend rapidement à différencier les échantillons de *Real_{Data}* des échantillons de *Synthetic_{Data}*, rendant l'apprentissage difficile pour le *Generator_{Network}*. L'effondrement de mode, quant à lui, peut se produire lorsque le *Generator_{Network}* apprend à créer un nombre restreint d'échantillons qui ne reflètent pas fidèlement la diversité de la distribution de données sous-jacente.

Pour remédier à l'instabilité, plusieurs approches ont été proposées, comme la modification de la fonction de perte GAN pour la rendre plus stable pendant l'entraînement. Par exemple, le GAN de Wasserstein (WGAN) (1) remplace la fonction de perte GAN originale par la distance de Wasserstein, qui peut produire un entraînement plus stable. Plusieurs stratégies ont été développées pour aborder l'effondrement de mode, telles que l'ajout de bruit à l'entrée du *Generator_{Network}*, l'utilisation de la correspondance de caractéristiques (30) ou l'utilisation d'architectures différentes pour le *Generator_{Network}* et le *Discriminator_{Network}*, telles que le CycleGAN (34).

5.2 Stabilisation de l'entraînement des GAN

Plusieurs stratégies pour stabiliser l'entraînement des GAN et aborder les problèmes susmentionnés ont été proposées. Une technique est la discrimination par mini-batch (30), qui consiste à ajouter des fonctionnalités supplémentaires au *Discriminator_{Network}* pour lui permettre de comparer plusieurs échantillons à la fois et de les différencier. Cela améliore la diversité des échantillons produits et aide à prévenir l'effondrement de mode. Une autre technique est la normalisation spectrale (33), qui consiste à normaliser les poids du *Discriminator_{Network}* pour assurer que la constante de Lipschitz du réseau est limitée. Cela aide à empêcher le *Discriminator_{Network}* de devenir trop puissant et stabilise le *Training_{Process}*. D'autres techniques consistent à utiliser différentes fonctions de perte, comme la fonction de perte des moindres carrés GAN (LSGAN) (28), pour améliorer la stabilité du *Training_{Process}*. De plus, des techniques de régularisation, telles que la décroissance pondérale et le dropout, peuvent également être

utilisées pour prévenir le surapprentissage et améliorer les performances de généralisation des modèles.

En conclusion, l'entraînement des GAN est une tâche difficile qui nécessite une considération minutieuse de plusieurs facteurs pour obtenir des résultats stables et de haute qualité. Les principaux défis de l'entraînement des GAN comprennent l'instabilité et l'effondrement de mode, qui peuvent être abordés à l'aide de diverses techniques, telles que la modification de la fonction de perte GAN, l'utilisation de différentes architectures et l'ajout de régularisation. Des recherches supplémentaires sont nécessaires pour développer des techniques plus efficaces pour l'entraînement des GAN et améliorer leurs performances dans diverses applications.

6 Applications des GANs

Les GAN ont suscité beaucoup d'intérêt récemment en raison de leur capacité à produire des *SyntheticData* de haute qualité qui correspondent étroitement aux *RealData*. Les GAN ont diverses applications dans différents domaines, notamment la vision par ordinateur, le traitement du langage naturel et les soins de santé.

6.1 Synthèse d'images

L'une des applications les plus célèbres des GAN est la synthèse d'images, où les GAN sont utilisés pour produire de nouvelles images similaires à un ensemble d'images d'entraînement donné. Les GAN peuvent créer des images hautement réalistes qui peuvent être utilisées à diverses fins, telles que dans les jeux vidéo, la réalité virtuelle et la création de *SyntheticData* pour l'entraînement des modèles d'apprentissage automatique (ML). Les avancées récentes dans la synthèse d'images basées sur les GAN ont conduit au développement de plusieurs nouvelles techniques, telles que les progressive GANs, styleGAN (10), et BigGAN (3). Les progressive GANs produisent des images haute résolution en augmentant progressivement la taille des images produites, tandis que styleGAN permet de contrôler différents aspects des images produites tels que le style, la pose et l'expression faciale. BigGAN est capable de générer des images de haute qualité jusqu'à 512x512 pixels.

6.2 Augmentation de données

Les GAN sont également adaptés à l'augmentation de données, qui consiste à créer des *SyntheticData* pour étendre la taille des *TrainingData*. En fournissant des *TrainingData* supplémentaires proches des *RealData*, l'augmentation de données à l'aide de GAN peut améliorer les performances des modèles ML. Cette approche a été appliquée avec succès dans divers domaines tels que la détection d'objets, la classification d'images et la reconnaissance vocale (31).

6.3 Transfert de style

De plus, les GAN peuvent être utilisés pour le transfert de style, qui est le processus de transfert du style d'une image à une autre. Cette méthode peut être utilisée pour

créer de nouvelles images en fusionnant le contenu d'une image avec le style d'une autre. Le transfert de style à l'aide de GAN a montré des résultats prometteurs dans divers domaines tels que la conception de mode, l'art et la photographie (4).

6.4 Applications émergentes

Les GAN sont également utilisés pour des applications émergentes telles que la synthèse de vidéos et la synthèse de texte en images. Dans la synthèse de vidéos, les GAN sont utilisés pour produire de nouveaux cadres vidéo similaires aux cadres existants. Cette technique peut être utilisée pour créer des vidéos de haute qualité avec moins d'efforts manuels. La synthèse de texte en images à l'aide de GAN implique la génération d'images à partir de descriptions textuelles. Cette approche présente des applications potentielles dans la conception de mode, la conception d'intérieur et d'autres domaines où la capacité de produire des images à partir de descriptions textuelles peut être utile (29).

TAB. 1 – *Applications des GANs*

Application	Description
Synthèse d'images	Les GAN produisent de nouvelles images similaires à un ensemble d'images d'entraînement donné. Les avancées récentes incluent les progressive GANs, styleGAN et BigGAN.
Augmentation de données	Les GAN créent des <i>SyntheticData</i> pour étendre la taille des <i>TrainingData</i> , améliorant les performances des modèles ML.
Transfert de style	Les GAN transfèrent le style d'une image à une autre, créant de nouvelles images en combinant le contenu avec le style.
Applications émergentes	Les GAN sont également utilisés pour la synthèse de vidéos et la synthèse de texte en images, avec des applications potentielles dans la conception de mode, la conception d'intérieur et d'autres domaines.

7 Évaluation des GANs

En raison de l'absence d'une fonction objective claire, l'évaluation de la performance des GAN est une tâche difficile. Les GAN produisent des *SyntheticData* en apprenant la distribution sous-jacente des *TrainingData*, et la qualité des données produites dépend de différents facteurs tels que l'architecture du *GeneratorNetwork* et du *DiscriminatorNetwork*, l'algorithme d'optimisation utilisé et le choix des hyperparamètres. Pour évaluer la performance des GAN, diverses mesures ont été définies, notamment IS⁴ et FID⁵. Basé sur la précision de classification d'un modèle Inception préalablement entraîné, l'IS évalue la diversité et la qualité des images générées. En

4. IS=Inception Score

5. FID= Fréchet Inception Distance

utilisant un modèle Inception déjà entraîné, le FID calcule la séparation dans l'espace des caractéristiques entre les distributions des images réelles et créées.

Bien que ces mesures aient été largement utilisées dans la recherche sur les GAN, elles présentent des limites. Par exemple, l'IS est connu pour favoriser les modèles qui produisent des images facilement classées par le modèle Inception, même si elles sont de faible qualité ou manquent de diversité. Le FID peut être sensible au bruit et aux artefacts d'image, et peut ne pas toujours être corrélé avec la qualité visuelle. De plus, les deux mesures nécessitent des modèles Inception pré-entraînés, qui peuvent ne pas être facilement disponibles ou ne peuvent pas être adaptés à tous les types de données. Il y a une recherche en cours pour développer de nouvelles mesures d'évaluation qui peuvent mieux capturer la performance des GAN. Certaines propositions récentes comprennent KID⁶, qui mesure la distance entre les distributions des caractéristiques extraites du modèle Inception, et LPIPS⁷, qui mesure la similarité perceptuelle entre les images basée sur les activations d'un NN profond pré-entraîné.

En conclusion, l'évaluation de la performance des GAN est un aspect important de la recherche sur les GAN, et diverses mesures ont été proposées à cette fin. Cependant, les mesures d'évaluation actuelles présentent des limites, et il est nécessaire de développer de nouvelles mesures qui peuvent mieux capturer la performance des GAN.

8 Défis et questions ouvertes

Les GAN ont démontré un énorme potentiel pour produire des photos, des films et d'autres formes de données réalistes. Cependant, il existe divers obstacles et préoccupations non résolues qui doivent être abordés afin d'améliorer la performance et l'utilisabilité des GAN. Parmi les obstacles et les questions non résolues, on peut citer :

1. Effondrement de mode : L'effondrement de mode est un problème courant dans les GAN, où le *Generator_{Network}* ne crée qu'un ensemble limité de sorties, en ignorant d'autres sorties possibles. Cela peut entraîner un manque de diversité dans les données générées. Une cause possible de l'effondrement de mode est que le *Discriminator_{Network}* est trop fort par rapport au *Generator_{Network}*, ce qui conduit à ce que le *Generator_{Network}* produise des échantillons similaires qui trompent le *Discriminator_{Network}*. Les chercheurs explorent différentes techniques pour résoudre l'effondrement de mode, telles que l'ajout de termes de régularisation à la *LossFunction* ou l'utilisation de méthodes d'entraînement alternatives. D'autres techniques comprennent la modification de l'architecture du *Generator_{Network}* et du *Discriminator_{Network}* ou l'utilisation de méthodes d'optimisation plus avancées.
2. Instabilité de l'entraînement : Il peut être difficile d'entraîner des GAN, et le *TrainingProcess* peut être instable, entraînant des oscillations ou une divergence dans la perte du *Generator_{Network}* et du *Discriminator_{Network}*. Cela peut rendre difficile l'obtention d'une bonne performance. Une cause possible de l'instabilité de l'entraînement est le déséquilibre entre le *Generator_{Network}*

6. KID=Kernel Inception Distance

7. LPIPS=Learned Perceptual Image Patch Similarity

et le *Discriminator_{Network}*, où l'un domine l'autre. Une autre cause est le problème du gradient qui disparaît, où les gradients de la *LossFunction* deviennent trop petits pour mettre à jour les paramètres. Les chercheurs étudient diverses approches pour améliorer la stabilité de l'entraînement des GAN, telles que l'ajustement du taux d'apprentissage, l'utilisation d'algorithmes d'optimisation différents ou l'ajout de bruit au *TrainingProcess*. Une autre approche consiste à utiliser des architectures plus avancées, telles que les GAN de Wasserstein ou les GAN à normalisation spectrale, qui ont montré être plus stables pendant l'entraînement.

3. Métriques d'évaluation : Il existe un manque de métriques d'évaluation largement acceptées pour les GAN, ce qui rend difficile la comparaison de différents modèles et l'évaluation de leur performance de manière objective. Certaines métriques d'évaluation proposées comprennent FID, qui évalue la distance entre la distribution des échantillons produits et la distribution des échantillons réels dans un espace de caractéristiques, et IS, qui mesure la diversité et la qualité des échantillons produits en fonction de leurs scores de classification par un classificateur pré-entraîné. Cependant, ces métriques ont des limites et peuvent ne pas capturer tous les aspects des données produites. Les chercheurs explorent des métriques d'évaluation alternatives et des méthodes pour mieux quantifier la performance des GAN.
4. Scalabilité : Les GAN peuvent être coûteux en termes de calcul pour l'entraînement et nécessitent des volumes énormes de données. Par conséquent, leur scalabilité et leur applicabilité aux défis du monde réel peuvent être limitées. L'apprentissage par transfert, qui consiste à affiner un modèle pré-entraîné sur un nouveau jeu de données, est une méthode pour améliorer la scalabilité des GAN. Une autre option consiste à tirer parti du calcul parallèle, également connu sous le nom d'entraînement distribué sur de nombreux GPU, ou à utiliser des ressources informatiques basées sur le cloud. Les chercheurs examinent également des techniques pour réduire la quantité de données nécessaires à l'entraînement des GAN, également connues sous le nom d'apprentissage semi-supervisé, ou pour enrichir les données d'entraînement avec des modèles génératifs.
5. Implications éthiques : Comme pour toute technologie, les GAN soulèvent des implications éthiques, en particulier dans le contexte de la génération d'images ou de vidéos réalistes. Les GAN peuvent être utilisés pour créer du contenu faux qui peut être utilisé à des fins malveillantes, telles que la diffusion de désinformation ou la création de deepfakes. Cela peut avoir des conséquences graves pour les individus et la société dans son ensemble. Les chercheurs et les décideurs politiques étudient des moyens d'atténuer ces risques et de promouvoir l'utilisation responsable des GAN, tels que le développement de méthodes de détection de deepfakes ou la création de lignes directrices pour l'utilisation éthique des données synthétiques.

9 Directions futures

Les GAN sont rapidement devenus l'un des domaines les plus passionnants de l'apprentissage profond depuis leur introduction en 2014. La capacité à générer des données réalistes à l'aide de GAN présente de nombreuses applications dans divers domaines tels que la vision par ordinateur, le traitement du langage naturel et la synthèse audio. Alors que les GAN continuent d'évoluer, les chercheurs et les praticiens peuvent explorer plusieurs directions futures pour améliorer leur évolutivité, leur stabilité et leurs performances. Ces directions futures comprennent :

1. Améliorer l'évolutivité des GAN : les GAN peuvent être coûteux en termes de calcul et nécessitent de grandes quantités de données pour l'entraînement. Les chercheurs explorent diverses techniques pour améliorer l'évolutivité des GAN, telles que l'utilisation de l'informatique parallèle, l'apprentissage par transfert ou la réduction de la quantité de données requise pour l'entraînement des GAN. Les praticiens peuvent expérimenter avec ces techniques pour améliorer l'évolutivité de leurs modèles GAN et les rendre plus applicables aux problèmes du monde réel.
2. Développer des architectures plus avancées : les architectures GAN ont considérablement évolué depuis leur création, de l'architecture GAN originale à des architectures plus avancées telles que les GAN de Wasserstein, les GAN progressifs et les StyleGAN. Les chercheurs peuvent continuer à explorer et à développer de nouvelles architectures qui sont plus stables, évolutives et capables de générer des données de haute qualité.
3. Incorporer la connaissance du domaine : les GAN peuvent bénéficier de l'incorporation de la connaissance du domaine, telle que les lois physiques ou la connaissance experte dans un domaine particulier. Les chercheurs peuvent explorer des moyens d'incorporer la connaissance du domaine dans les modèles GAN pour améliorer leurs performances et leurs capacités de généralisation.
4. Explorer de nouvelles applications des GAN : les GAN ont déjà été appliqués à une large gamme de domaines, tels que la vision par ordinateur, le traitement du langage naturel et la synthèse audio. Cependant, il existe encore de nombreuses applications potentielles des GAN qui n'ont pas encore été explorées, telles que dans les domaines de la santé, de la finance ou des sciences sociales. Les chercheurs peuvent explorer de nouveaux cas d'utilisation et applications des GAN dans ces domaines, et les praticiens peuvent expérimenter l'application des GAN à de nouveaux problèmes dans leur domaine.
5. Développer des métriques d'évaluation plus robustes : il est nécessaire de disposer de métriques d'évaluation plus robustes pour les GAN qui peuvent capturer tous les aspects des données produites, tels que la qualité visuelle, la diversité et le réalisme. Les chercheurs peuvent explorer de nouvelles métriques et méthodes d'évaluation plus fiables et objectives, et les praticiens peuvent utiliser ces métriques pour estimer les performances de leurs modèles GAN.
6. Utilisation de méthodes formelles : une autre direction future pour les GAN est l'utilisation de méthodes formelles pour améliorer leur fiabilité et leur sécurité

(13; 18; 25). Les méthodes formelles sont un ensemble de techniques et d'outils mathématiques utilisés pour analyser et vérifier rigoureusement les systèmes logiciels et matériels (23; 21; 26). L'utilisation de méthodes formelles dans les GAN peut aider à garantir qu'ils produisent des sorties qui répondent à certaines exigences de sécurité et de fiabilité (19; 20; 14). Les méthodes formelles peuvent être utilisées pour vérifier les propriétés des GAN, telles que l'absence de certains types d'erreurs ou la correction de certaines opérations (27; 11). Par exemple, les méthodes formelles peuvent être utilisées pour vérifier que les données produites ne violent pas certaines contraintes de sécurité, ou que le processus d'entraînement ne diverge pas ou ne présente pas de comportement indésirable (15; 24).

Comme résumé dans le Tableau 2, il y a de nombreuses directions prometteuses pour les GAN, notamment l'amélioration de leur évolutivité et de leur stabilité, le développement d'architectures plus avancées, l'incorporation de la connaissance du domaine, l'exploration de nouvelles applications des GAN et le développement de métriques d'évaluation plus robustes. En continuant à repousser les limites de la recherche et du développement des GAN, les chercheurs et les praticiens peuvent exploiter tout le potentiel des GAN et stimuler l'innovation dans un large éventail de domaines.

10 Conclusion

Les GAN sont devenus un outil puissant pour produire des données réalistes dans une variété de disciplines. Dans cet article, nous avons proposé un guide complet des GAN, couvrant leur architecture, les fonctions de perte, les méthodes d'entraînement, les applications, les métriques d'évaluation, les défis et les directions futures. Nous avons examiné le développement historique des GAN, de leur formulation originale aux avancées plus récentes telles que les GAN de Wasserstein et les StyleGAN. Nous avons discuté des choix de conception clés et des variations dans l'architecture des GAN, ainsi que des différentes fonctions de perte utilisées pour former des modèles GAN. Nous avons également exploré les diverses applications des GAN, de la synthèse d'images à la NLP et à la synthèse audio, et examiné les métriques d'évaluation utilisées pour évaluer la diversité et la qualité des données produites par les GAN. De plus, nous avons souligné les défis et les problèmes ouverts dans la recherche sur les GAN, tels que l'instabilité de l'entraînement, l'effondrement du mode et les considérations éthiques. Enfin, nous avons donné un aperçu des directions futures de la recherche sur les GAN, notamment l'amélioration de l'évolutivité, le développement de nouvelles architectures, l'incorporation de la connaissance du domaine et l'exploration de nouvelles applications.

Dans l'ensemble, les GAN représentent un domaine de recherche en rapide évolution avec un potentiel considérable pour l'innovation et l'impact. En fournissant un guide complet des GAN, nous espérons faciliter davantage la recherche et le développement dans ce domaine et inspirer de nouvelles applications et cas d'utilisation. Nous anticipons que, à mesure que les GAN s'amélioreront, ils joueront un rôle de plus en plus important dans la production et la synthèse de données, ouvrant de nouvelles voies pour la découverte scientifique et l'expression artistique.

TAB. 2 – Directions futures des GAN

Direction	Description
Amélioration de l'évolutivité	Les chercheurs explorent diverses techniques pour améliorer l'évolutivité des GAN, telles que l'utilisation de l'informatique parallèle, l'apprentissage par transfert ou la réduction de la quantité de données requise pour l'entraînement des GAN.
Développement d'architectures avancées	Les chercheurs peuvent continuer à explorer et à développer de nouvelles architectures qui sont plus stables, évolutives et capables de générer des données de haute qualité.
Incorporation de la connaissance du domaine	Les GAN peuvent bénéficier de l'incorporation de la connaissance du domaine, telle que les lois physiques ou la connaissance experte dans un domaine particulier. Les chercheurs peuvent explorer des moyens d'incorporer la connaissance du domaine dans les modèles GAN.
Exploration de nouvelles applications	Les chercheurs peuvent explorer de nouveaux cas d'utilisation et applications des GAN dans des domaines tels que la santé, la finance ou les sciences sociales. Les praticiens peuvent expérimenter l'application des GAN à de nouveaux problèmes dans leur domaine.
Développement de métriques d'évaluation robustes	Les chercheurs peuvent explorer de nouvelles métriques et méthodes d'évaluation plus fiables et objectives, et les praticiens peuvent utiliser ces métriques pour estimer les performances de leurs modèles GAN.
Utilisation de méthodes formelles	Les méthodes formelles peuvent être utilisées pour améliorer la fiabilité et la sécurité des GAN en vérifiant les propriétés des GAN ou en générant des exemples adversaires pour évaluer leur robustesse (12; 17).

Références

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [2] Wadii Boulila, Maha Driss, Eman Alshantqi, Mohamed Al-Sarem, Faisal Saeed, and Moez Krichen. Weight initialization techniques for deep learning algorithms in remote sensing : Recent trends and future perspectives. *Advances on Smart and Soft Computing : Proceedings of ICACIn 2021*, pages 477–484, 2022.
- [3] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv :1809.11096*, 2018.
- [4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.
- [5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [6] Jon Gauthier. Conditional generative adversarial nets for convolutional face generation. *Class project for Stanford CS231N : convolutional neural networks for visual recognition, Winter semester, 2014(5) :2*, 2014.
- [7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11) :139–144, 2020.
- [9] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [10] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [11] Moez Krichen. *Model-based testing for real-time systems*. PhD thesis, PhD thesis, PhD thesis, Universit Joseph Fourier (December 2007), 2007.
- [12] Moez Krichen. A formal framework for conformance testing of distributed real-time systems. In *International Conference On Principles Of Distributed Systems*, pages 139–142. Springer, 2010.
- [13] Moez Krichen. *Contributions to model-based testing of dynamic and distributed real-time systems*. PhD thesis, École Nationale d’Ingénieurs de Sfax (Tunisie), 2018.
- [14] Moez Krichen, Afef Jmal Maâlej, and Mariam Lahami. A model-based approach to combine conformance and load tests : an ehealth case study. *International Journal of Critical Computer-Based Systems*, 8(3-4) :282–310, 2018.

- [15] Moez Krichen, Seifeddine Mechti, Roobaea Alroobaea, Elyes Said, Parminder Singh, Osamah Ibrahim Khalaf, and Mehedi Masud. A formal testing model for operating room control system using internet of things. *Computers, Materials & Continua*, 66(3) :2997–3011, 2021.
- [16] Moez Krichen, Alaeddine Mihoub, Mohammed Y Alzahrani, Wilfried Yves Hamilton Adoni, and Tarik Nahhal. Are formal methods applicable to machine learning and artificial intelligence? In *2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, pages 48–53. IEEE, 2022.
- [17] Moez Krichen and Stavros Tripakis. State identification problems for timed automata. In *TestCom*, volume 5, pages 175–191, 2005.
- [18] Moez Krichen and Stavros Tripakis. Interesting properties of the real-time conformance relation tioco. In *Theoretical Aspects of Computing-ICTAC 2006 : Third International Colloquium, Tunis, Tunisia, November 20-24, 2006. Proceedings 3*, pages 317–331. Springer Berlin Heidelberg, 2006.
- [19] Mariam Lahami, Moez Krichen, Hajer Barhoumi, and Mohamed Jmaiel. Selective test generation approach for testing dynamic behavioral adaptations. In *Testing Software and Systems : 27th IFIP WG 6.1 International Conference, ICTSS 2015, Sharjah and Dubai, United Arab Emirates, November 23-25, 2015, Proceedings 27*, pages 224–239. Springer International Publishing, 2015.
- [20] Mariam Lahami, Moez Krichen, Mariam Bouchakwa, and Mohamed Jmaiel. Using knapsack problem model to design a resource aware test architecture for adaptable and distributed systems. In *Testing Software and Systems : 24th IFIP WG 6.1 International Conference, ICTSS 2012, Aalborg, Denmark, November 19-21, 2012. Proceedings 24*, pages 103–118. Springer Berlin Heidelberg, 2012.
- [21] Mariam Lahami, Moez Krichen, and Mohamed Jmaïel. Runtime testing approach of structural adaptations for dynamic and distributed systems. *International Journal of Computer Applications in Technology*, 51(4) :259–272, 2015.
- [22] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553) :436–444, 2015.
- [23] Afef Jmal Maâlej, Manel Hamza, Moez Krichen, and Mohamed Jmaiel. Automated significant load testing for ws-bpel compositions. In *2013 IEEE sixth international conference on software testing, verification and validation workshops*, pages 144–153. IEEE, 2013.
- [24] Afef Jmal Maâlej and Moez Krichen. A model based approach to combine load and functional tests for service oriented architectures. In *VECoS*, pages 123–140, 2016.
- [25] Afef Jmal Maâlej, Moez Krichen, and Mohamed Jmaiel. Conformance testing of ws-bpel compositions under various load conditions. In *2012 IEEE 36th annual computer software and applications conference*, pages 371–371. IEEE, 2012.
- [26] Afef Jmal Maâlej, Moez Krichen, and Mohamed Jmaiel. Model-based conformance testing of ws-bpel compositions. In *2012 IEEE 36th annual computer software and applications conference workshops*, pages 452–457. IEEE, 2012.

- [27] Afef Jmal Maâlej, Mariam Lahami, Moez Krichen, and Mohamed Jmaïel. Distributed and resource-aware load testing of ws-bpel compositions. In *ICEIS (2)*, pages 29–38, 2018.
- [28] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.
- [29] Scott Reed, Zeynep Akata, Xinchun Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *International conference on machine learning*, pages 1060–1069. PMLR, 2016.
- [30] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.
- [31] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1) :1–48, 2019.
- [32] Xuan Xia, Xizhou Pan, Nan Li, Xing He, Lin Ma, Xiaoguang Zhang, and Ning Ding. Gan-based anomaly detection : a review. *Neurocomputing*, 2022.
- [33] Chao Xiaopeng, Cao Jiangzhong, Lu Yuqin, and Dai Qingyun. Improved training of spectral normalization generative adversarial networks. In *2020 2nd World Symposium on Artificial Intelligence (WSAI)*, pages 24–28. IEEE, 2020.
- [34] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.