

1st workshop on Advances in Deep Learning for Image and Immersive Technologies (ADL2IT)

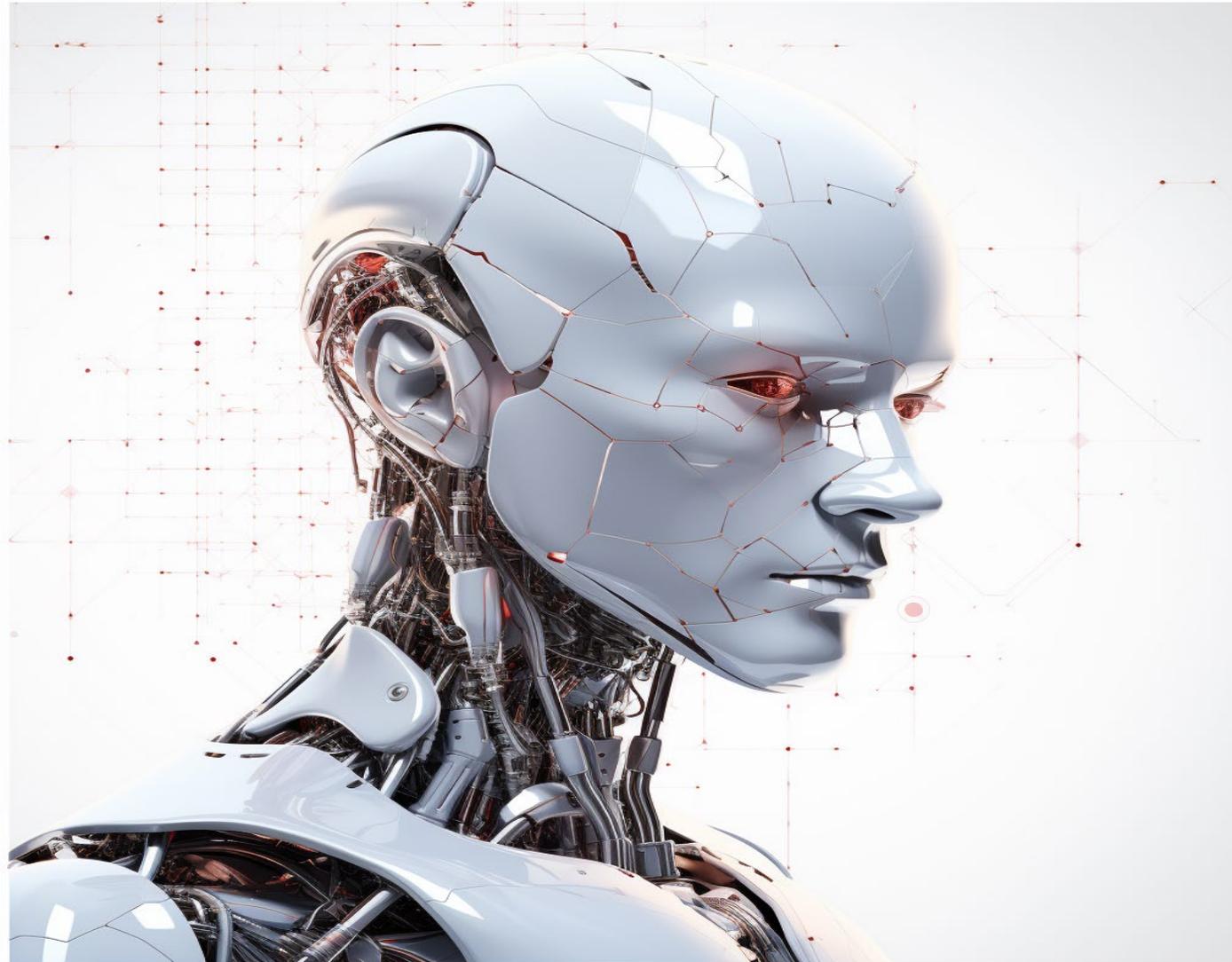
10 - 12 October 2023 Biskra, Algeria



RENDU NEURONAL, DES MODÈLES GÉNÉRATIFS AUX MODÈLES DE DIFFUSION

1st workshop on Advances in Deep Learning for Image and Immersive Technologies (ADL2IT)

10 - 12 October 2023 Biskra, Algeria



PLAN

- Introduction
- Rendu neuronal
- Réseaux Adversaire Génératif
- Application des GAN
- Modèles de diffusion
- Application des Modèles de diffusion

INTRODUCTION



RENDU CLASSIQUE

- Chaque objet est représenté par un maillage polygonal. Ce maillage polygonal peut être soit modélisé par un artiste, soit numérisé.
- Ces deux processus sont manuels et prennent du temps.
- Chaque surface possède également un matériau qui détermine son apparence.
- Créer une scène 3D est un processus compliqué et complexe.



INTRODUCTION



QUESTION

- Que se passerait-il si nous n'avions pas besoin de spécifier chaque objet et de calculer chaque réflexion de lumière?

RÉPONSE

- les progrès de la vision par ordinateur et de l'apprentissage automatique ont donné naissance à de nouvelles approches de synthèse d'images, dites
- **Les approches génératifs.**

INTRODUCTION



RENDU CLASSIQUE



- plusieurs jours

MODÈLE GÉNÉRATIF



- plusieurs secondes

INTRODUCTION



RENDU CLASSIQUE



- plusieurs heures

MODÈLE GÉNÉRATIF



- plusieurs secondes



RENDU CLASSIQUE +

- Produire des images photoréalistes.
- Rendu physiquement précis.
- Matériaux complexes.
- Éclairage précis.
- Simuler des effets optiques complexes

MODÈLE GÉNÉRATIF +

- Peut générer des images en temps réel.
- Capable de produire des images de la qualité photoréalistes.
- Adapter aux changements dynamiques des scènes.
- Optimisé pour l'accélération GPU.



RENDU CLASSIQUE -

- Nécessitent des calculs importants et des ressources.
- Moins adaptable aux changements rapides de scènes.
- Les limitations de performances peuvent avoir un impact l'interactivité dans des applications comme la RV,RA.

MODÈLE GÉNÉRATIF -

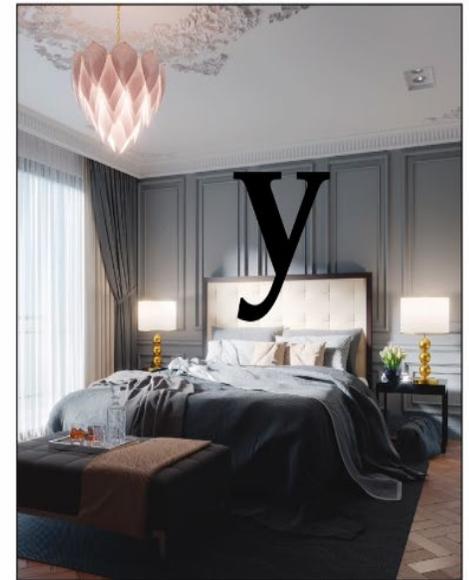
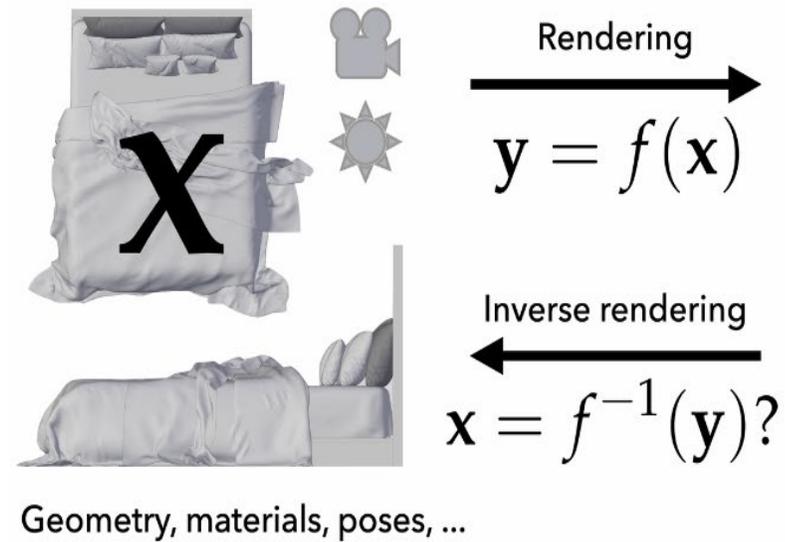
- Temps d'entraînement.
- Exigences matérielles.
- Artefacts et images irréalistes.
- Modèles complexes.

Le rendu neuronal comme solution de secours

LE RENDU NEURONAL



- Le rendu neuronal combine l'apprentissage automatique génératif avec des connaissances physiques issues de l'infographie.



WENZEL JAKOB, PROCEEDINGS OF SIGGRAPH 2022

RÉSEAUX ADVERSES GÉNÉRATIFS (GAN)



GAN

- C'est un réseau de neurones qui appartient à l'ensemble de l'IA générative.
- Ils sont capables de générer de nouveaux contenu qui n'a jamais été généré auparavant.
- L'architecture se compose de deux réseaux de neurones : le générateur et le discriminateur.

Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie*, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair[†], Aaron Courville, Yoshua Bengio[‡]
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G . The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D , a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

- <https://arxiv.org/abs/1406.2661>

RÉSEAUX ADVERSES GÉNÉRATIFS (GAN)



- Deux éléments principaux composent l'architecture de GAN: le générateur et le discriminateur :
- **Le générateur G** est chargé de modéliser la distribution des données.
- **Le discriminateur D** évalue la probabilité qu'un échantillon donné provienne du données d'entraînement plutôt que d'être générées par G.

RÉSEAUX ADVERSES GÉNÉRATIFS (GAN)



$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))]$$

- discriminator loss.
- z is a noise distribution.
- generator loss.

RÉSEAUX ADVERSES GÉNÉRATIFS (GAN)

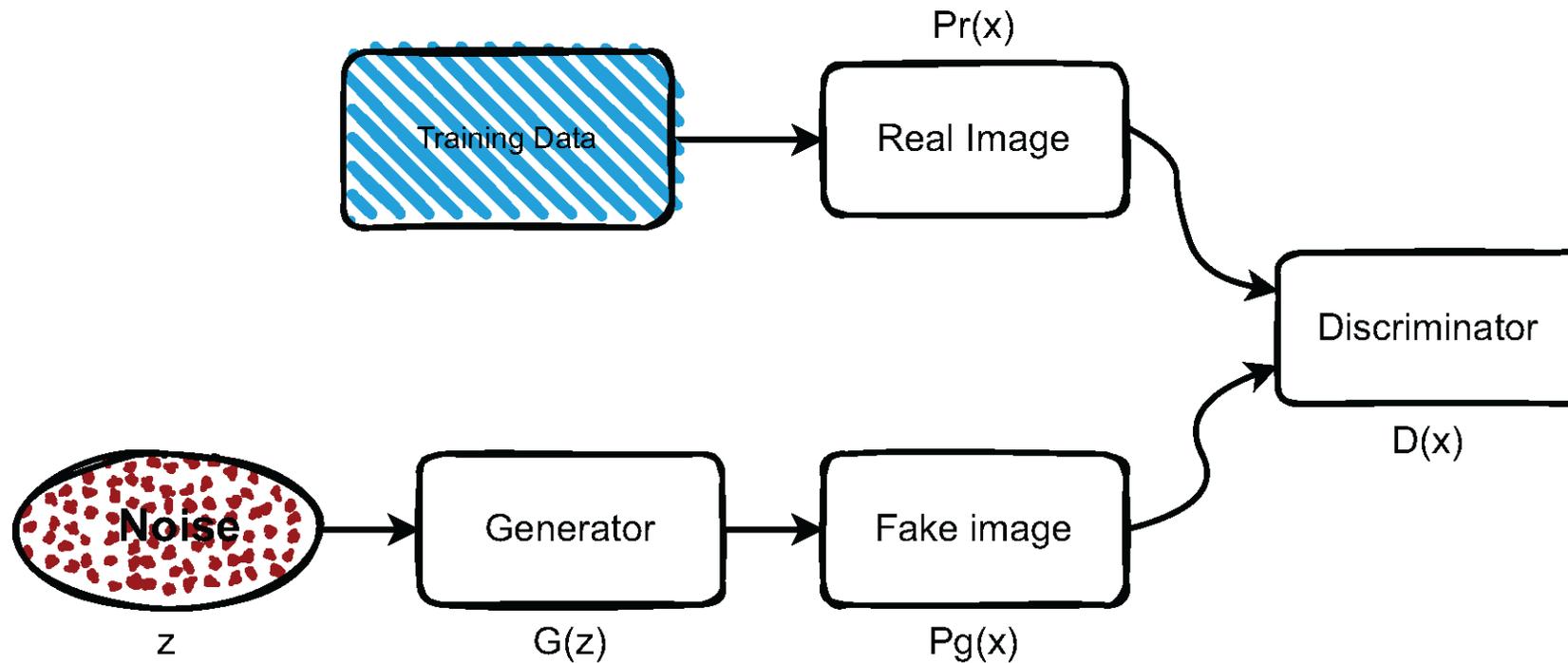


$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))]$$

- discriminator loss.
- z is a noise distribution.
- generator loss.

- $P_{data}(x)$ représente la distribution réelle des données.
- $D(x)$ représente la probabilité que x provienne des données.
- $D(x)$ est formé pour maximiser la probabilité d'attribuer la bonne étiquette aux deux formations exemples et échantillons de G . G est formé pour minimiser $\log(1 - D(G(z)))$.

RÉSEAUX ADVERSES GÉNÉRATIFS (GAN)



APPLICATIONS GAN



GÉNÉREZ DES VISAGES DE PERSONNES QUI N'EXISTENT PAS

- [https : //www.thispersondoesnotexist.com/](https://www.thispersondoesnotexist.com/)



APPLICATIONS GAN

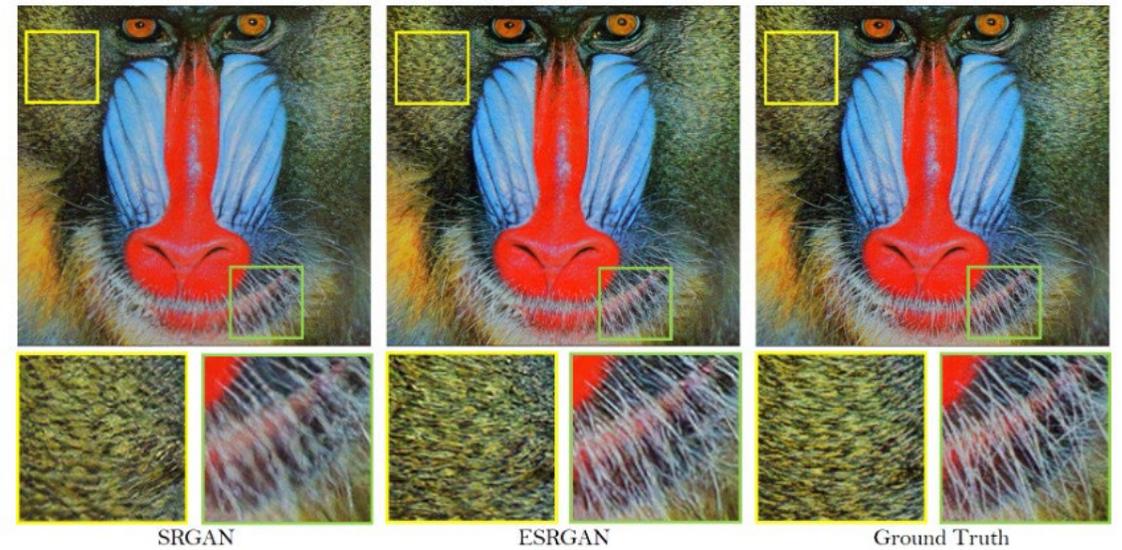


AUGMENTATION DES DONNÉES



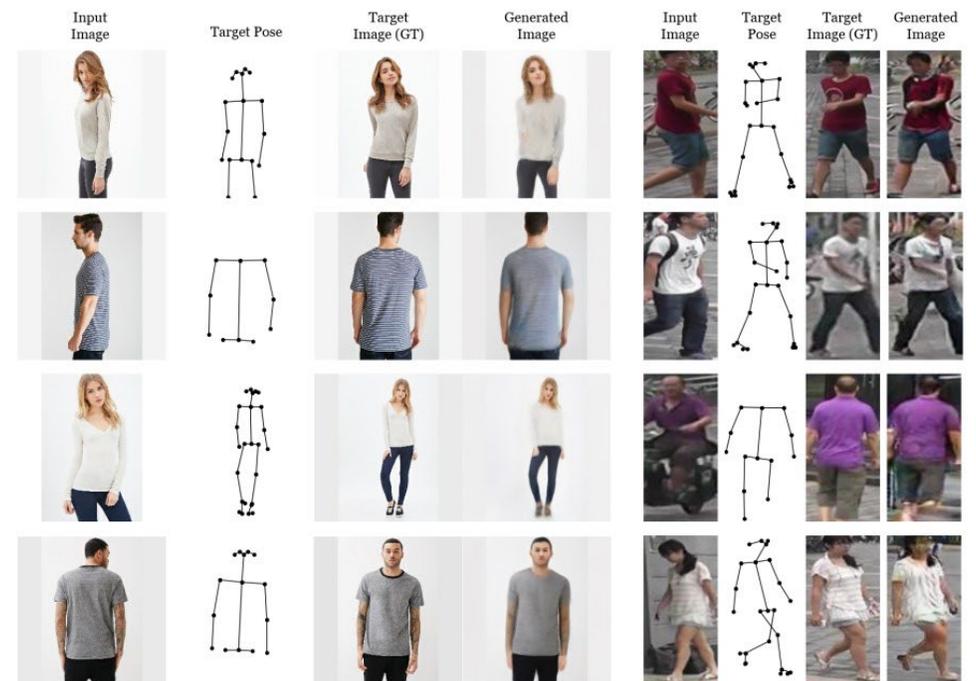
- <https://arxiv.org/abs/1809.11096>

SURÉCHANTILLONNER UNE IMAGE BASSE RÉSOLUTION EN UNE RÉSOLUTION PLUS ÉLEVÉE



- ESRGAN paper

ESTIMATION DE POSE



- <https://arxiv.org/pdf/2001.01259v1.pdf>



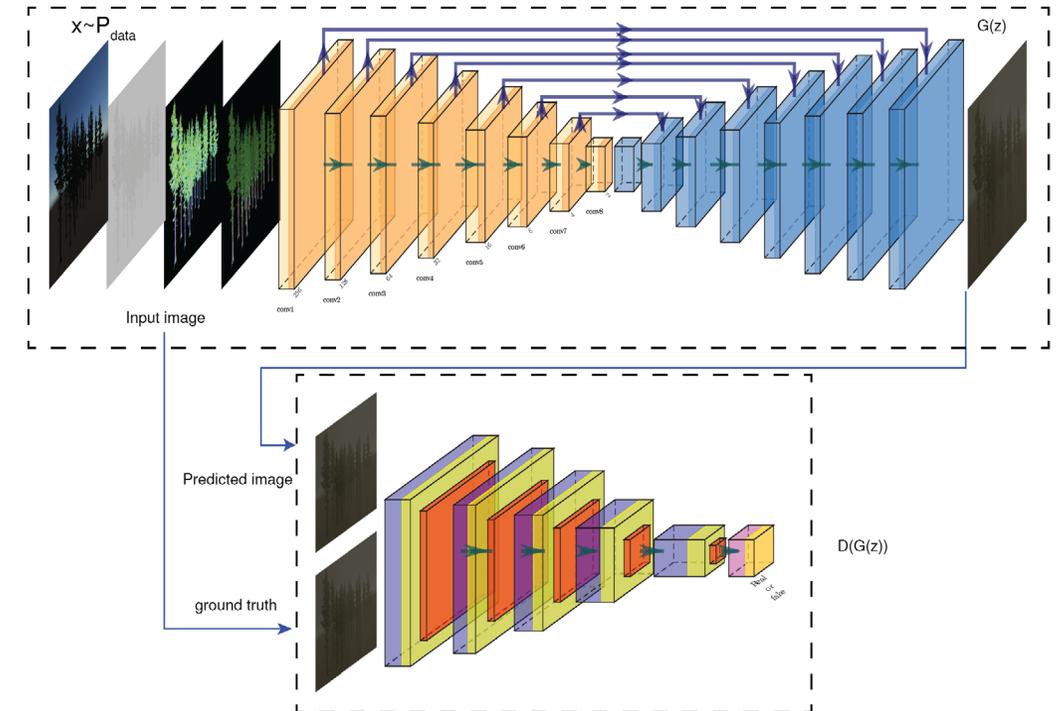
NOS CONTRIBUTIONS

LESIA



FOREST FOG RENDERING USING GENERATIVE ADVERSARIAL NETWORKS

- **L'objectif :**
- Estimez le brouillard dans les scènes forestières.
- Solution indépendante de la géométrie tridimensionnelle explicite de la scène.
- Nouvelles techniques basées sur le GAN.
- La solution ne nécessite aucun traitement de l'image de sortie.



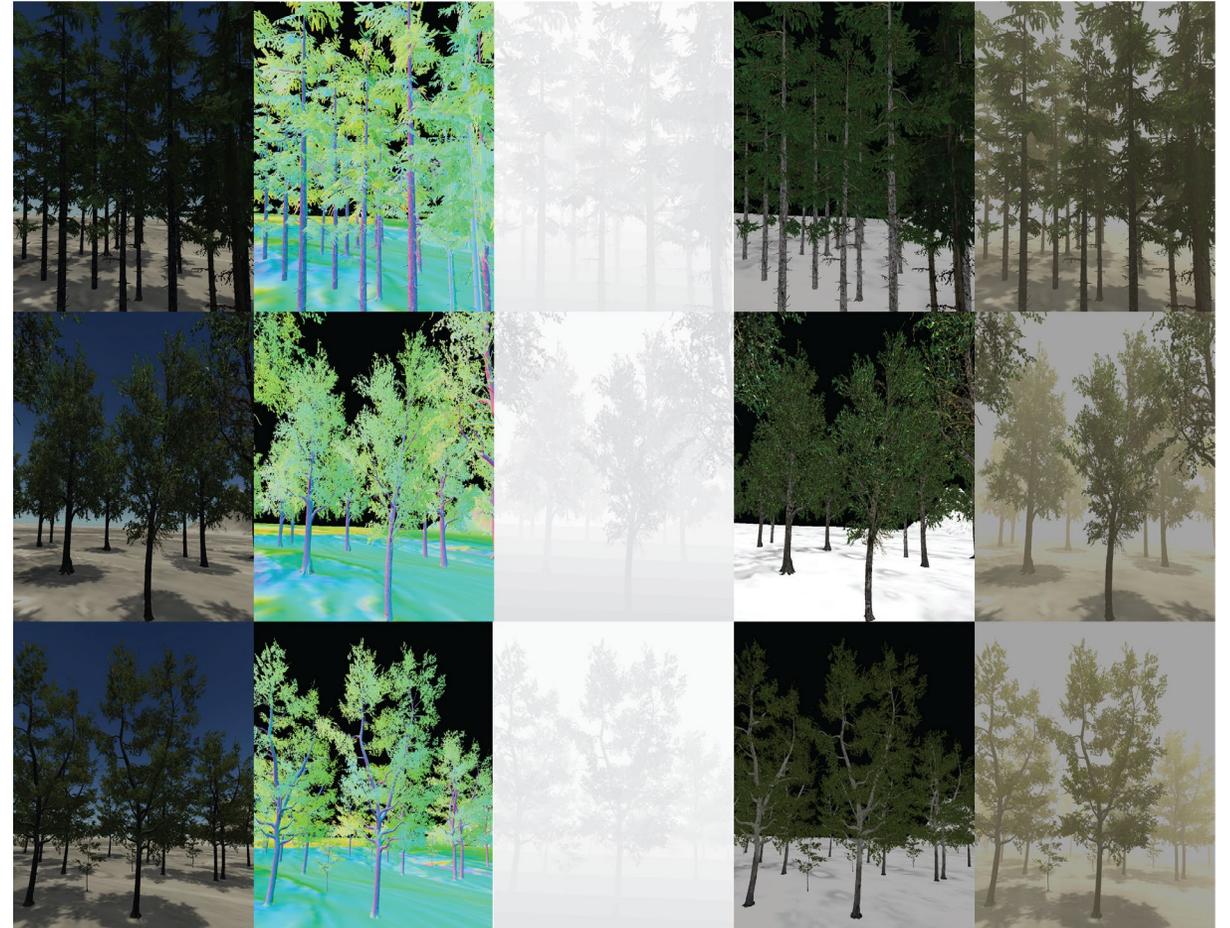
Abbas, F., & Babahenini, M. C. (2023). Forest fog rendering using generative adversarial networks. *The Visual Computer*, 39(3), 943-952.

CONTRIBUTION 1



COLLECTE DES DONNÉES

- Images de notre jeu de données synthétiques.
- RGB map sans brouillard.
- Normal map.
- Depth map.
- Albedo map.
- RGB map avec brouillard.

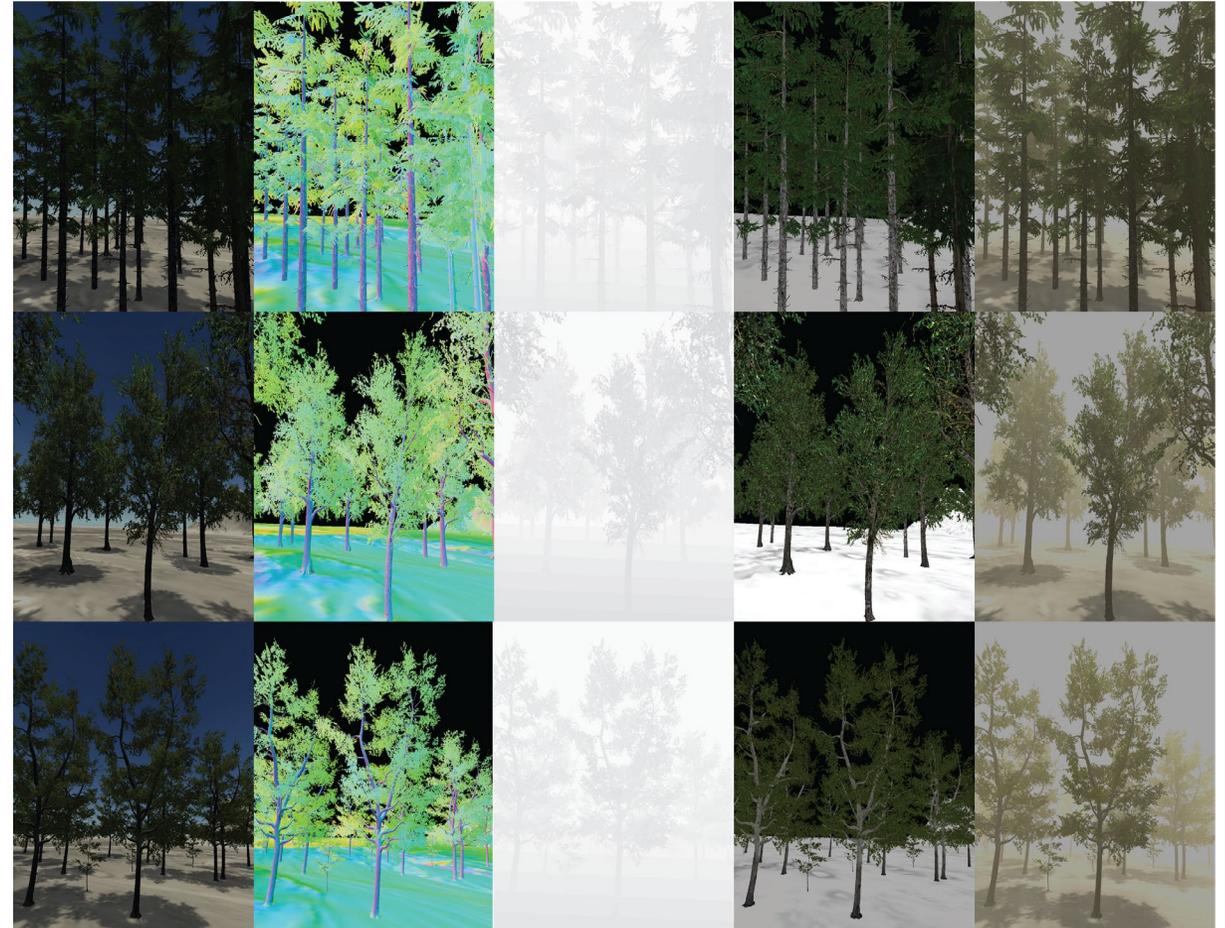


CONTRIBUTION 1



COLLECTE DES DONNÉES

- Pour chaque scène nous varions le nombre d'arbres et les texture des feuilles ainsi la texture du tronc.
- Nous effectuons des rotations autour de l'axe y.
- Pour chaque angle nous récupérons depuis le Gbuffer les cartes : RGB map sans brouillard, Normal map, Depth map, Albedo map, RGB map avec brouillard.



CONTRIBUTION 1



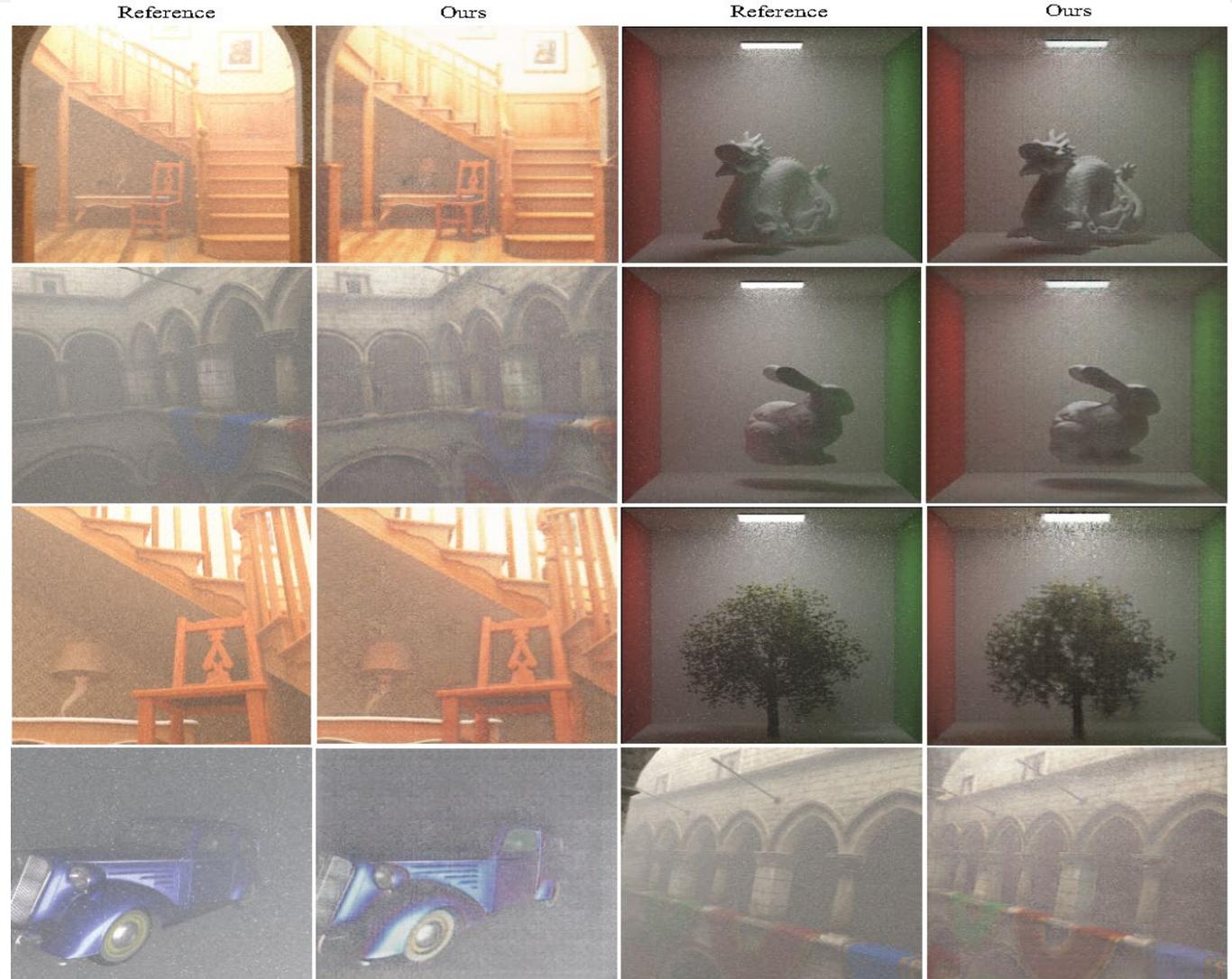
Résultat d'une forêts avec
brouillard généré par notre
approche



CONTRIBUTION 1



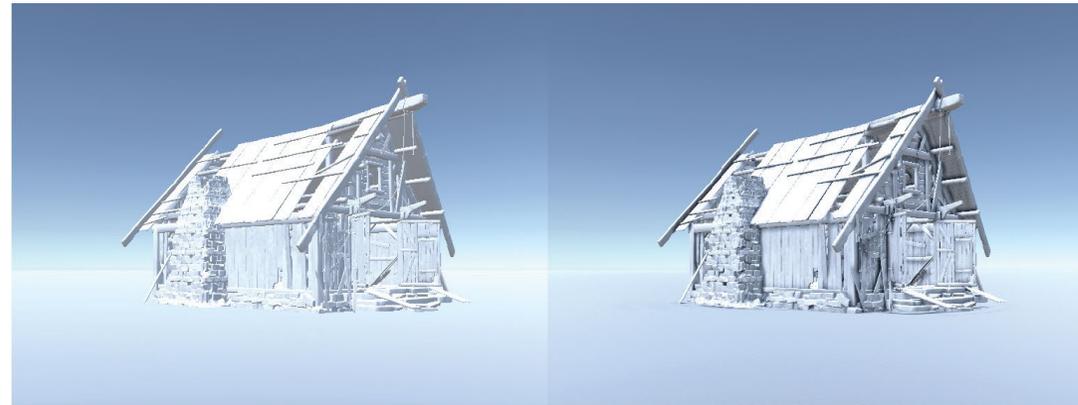
- Les images de références sont rendues avec le **volume path tracing** , en utilisant un milieu participant homogène et une fonction de phase isotrope.
- Les coefficients de diffusion $\sigma_s = 0,95$.
- les coefficients d'absorption $\sigma_a = 0,0$.
- Notre modèle peut être généralisé à d'autres scènes complexes et donne des résultats similaires à l'image de référence.



CONTRIBUTION 2



ATTENTIONAL CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR AMBIENT OCCLUSION APPROXIMATION



without AO

with AO

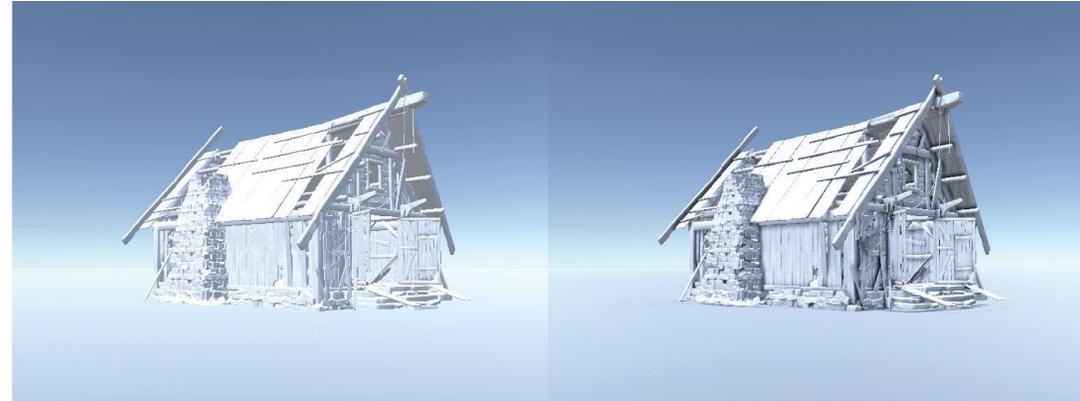


CONTRIBUTION 2



L'objectif :

- Une bonne approximation de l'Occultation Ambiante.
- Solution indépendante de la structure 3D des objets composant la scène.
- Un effet réaliste d'ombre douce.



without AO

with AO





Mécanisme d'attention

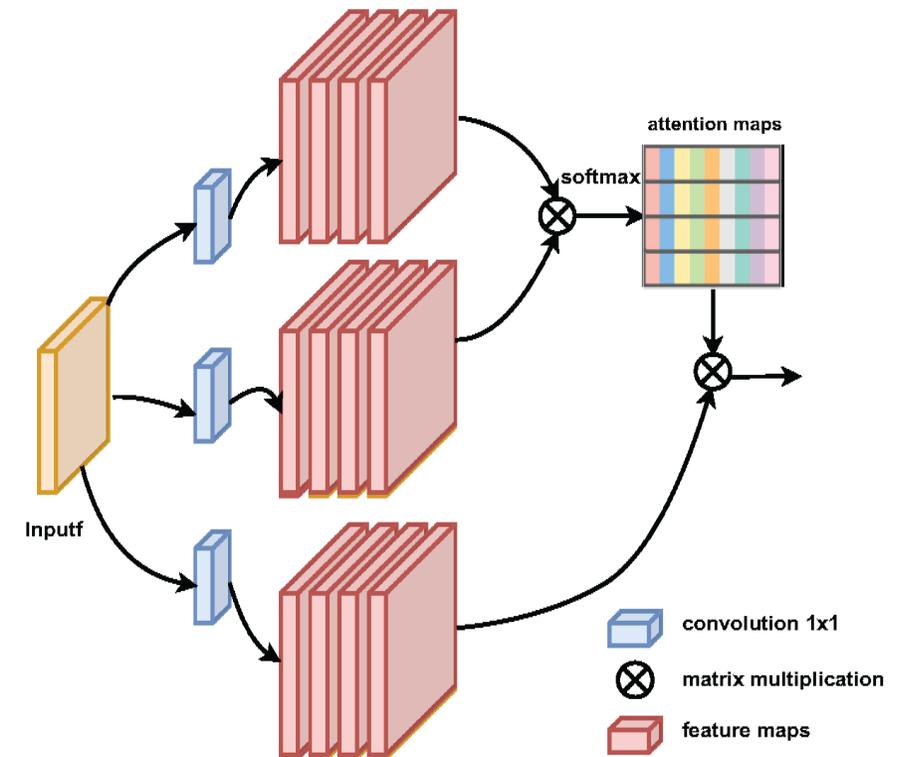
- Les modèles GAN ont la limitation qu'ils ne parviennent pas à apprendre les longues dépendances spatiales.

d'où la nécessité
d'utiliser un mécanisme
d'attention

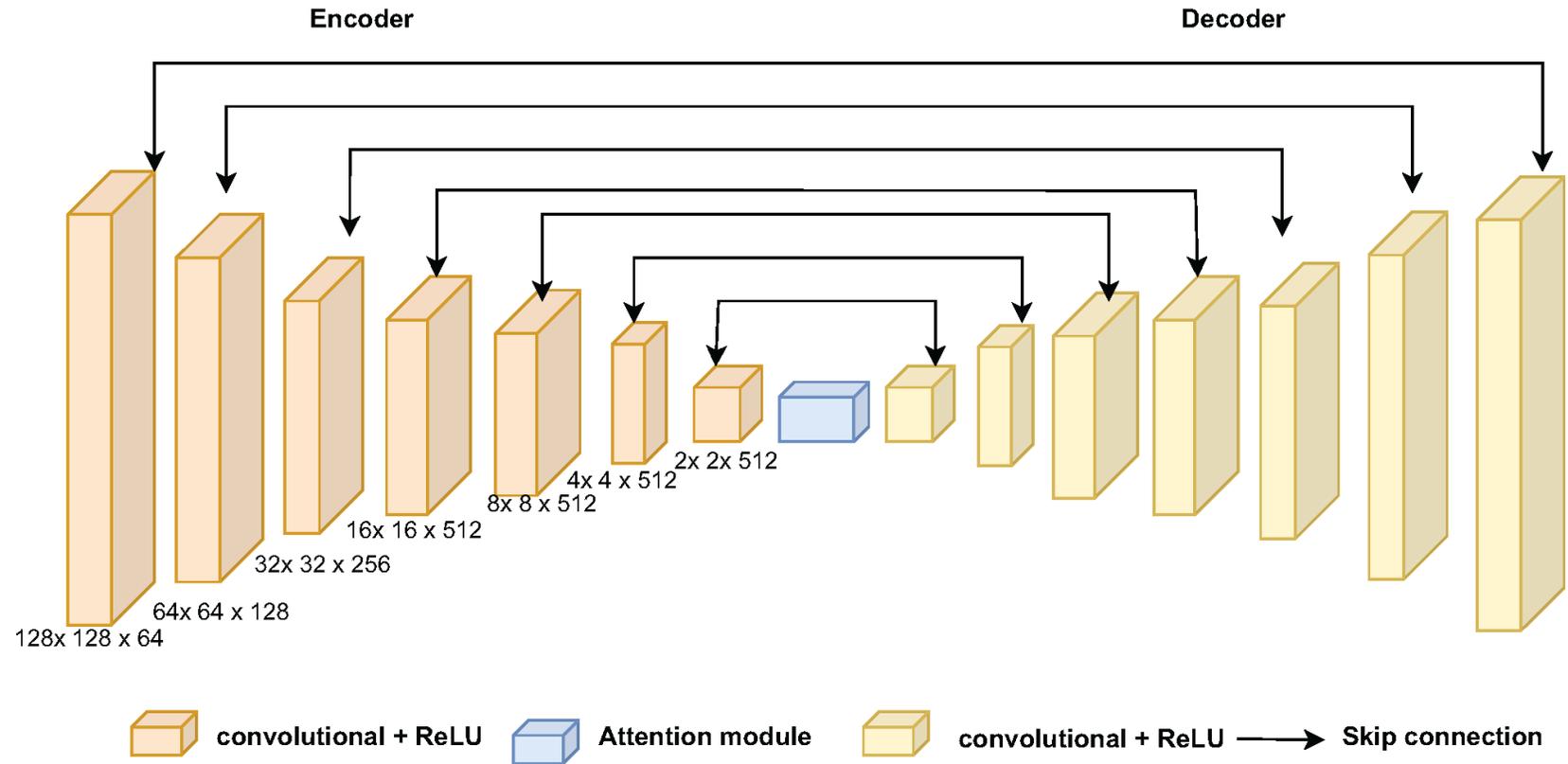
CONTRIBUTION 2



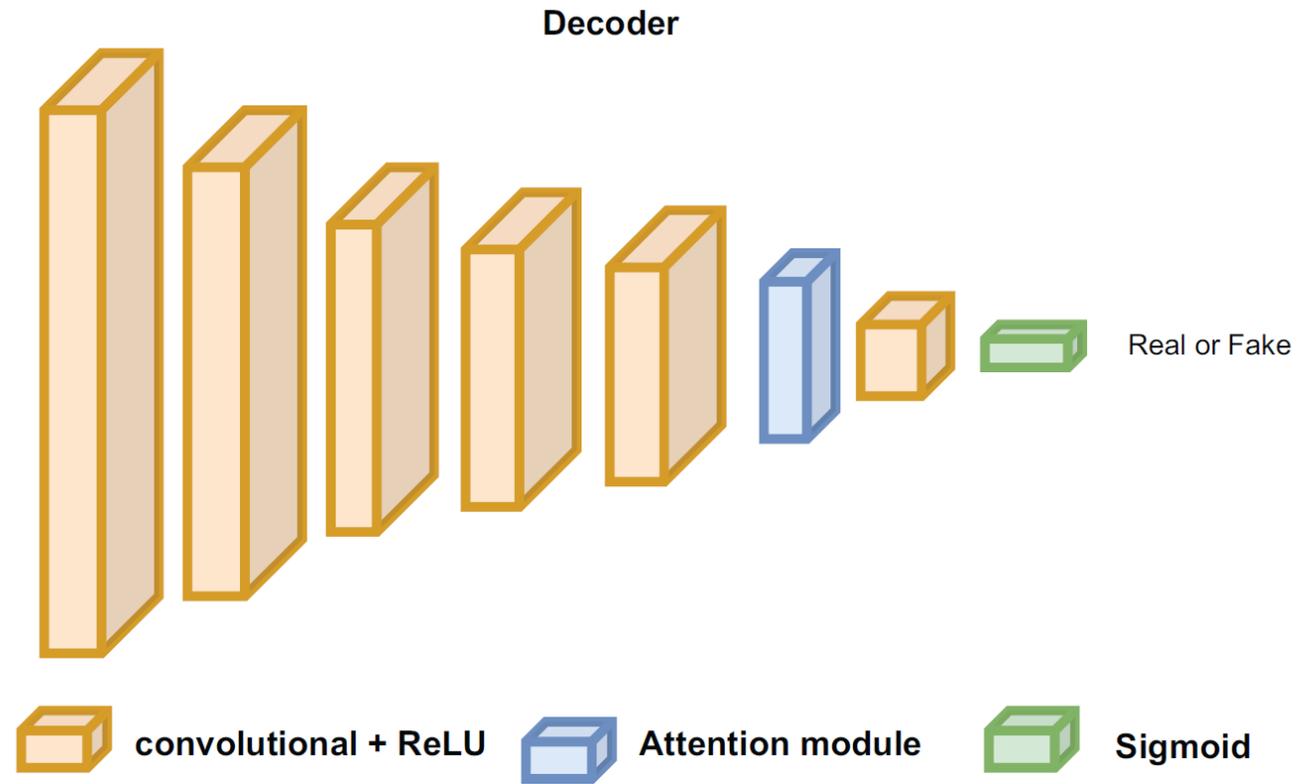
- Introduire un mécanisme d'attention dans le générateur et le discriminateur.
- Permettre au réseau de se concentrer sur les parties pertinentes difficiles à atteindre par une simple architecture.



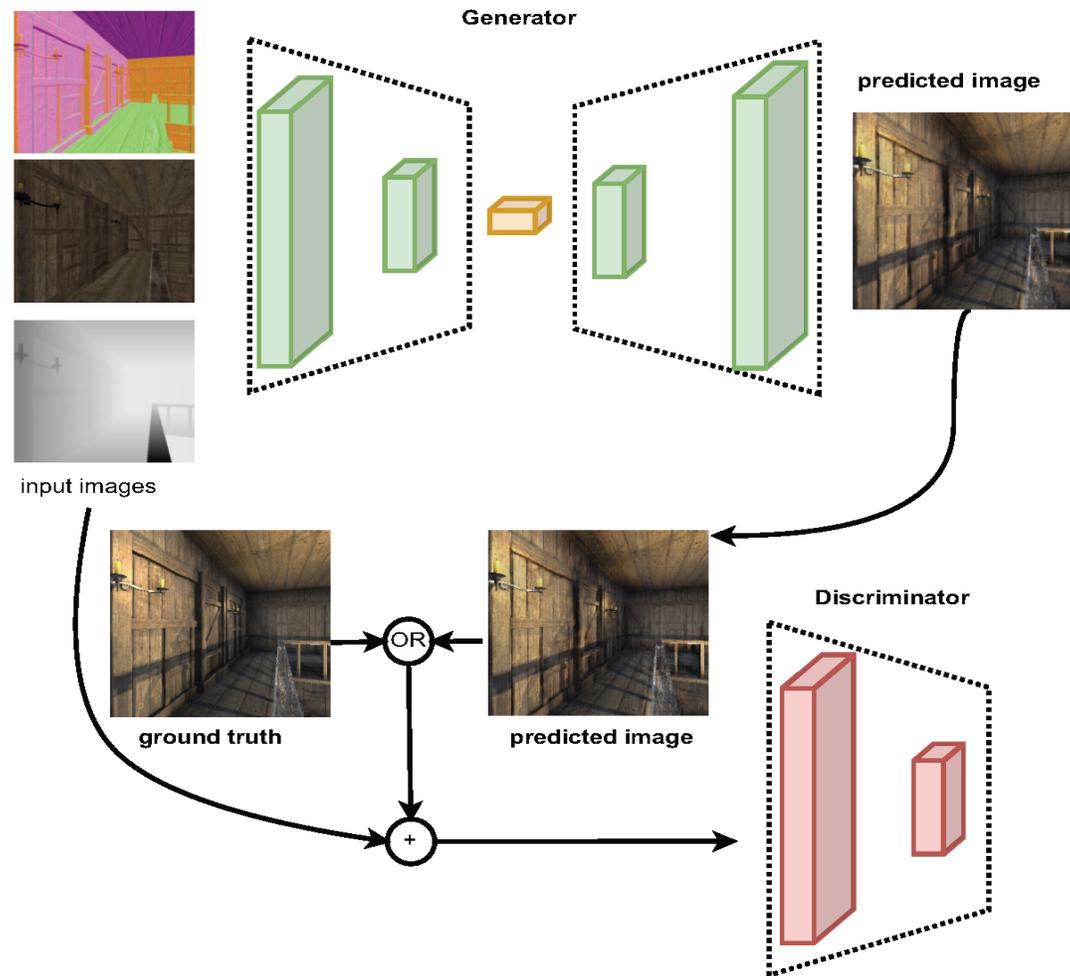
CONTRIBUTION 2



CONTRIBUTION 2



CONTRIBUTION 2



CONTRIBUTION 2



- Nous introduisons l'indice de similarité structurelle SSIM [Wang et al, 2004] dans la formule de la fonction de coût.
- Notre objectif final est donné par la formule suivante :

$$\min_G \min_D V(D, G) = I_{GAN}(G, D) + \beta I_{ssim}(I_{reference}, I_{predicted})$$

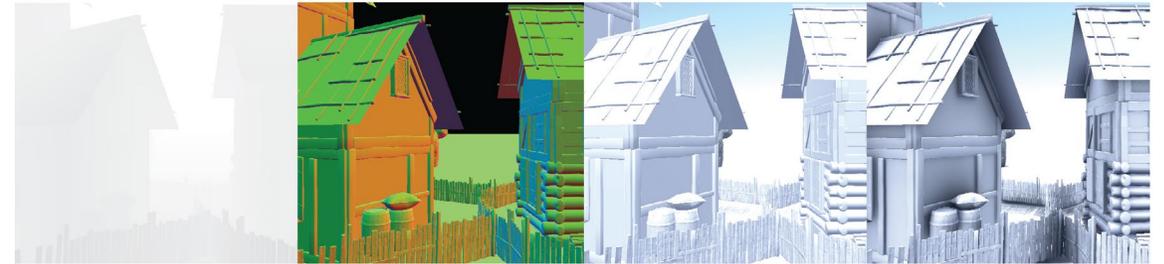
CONTRIBUTION 2



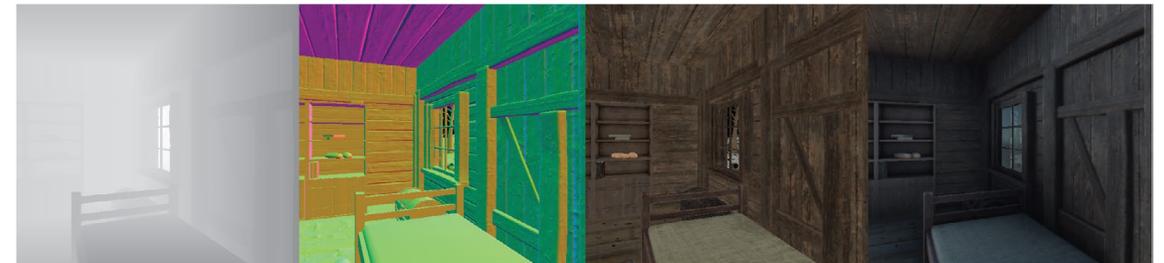
DATASET

Deux jeux de données :

- le premier : images de référence avec AO sans texture.
- le second : AO avec texture.
- 20 scènes pour la formation et 5 scènes pour le test.
- Les images générées sont d'une résolution de 256×256 pixels.
- Les images sont prises en tournant sur l'axes Y .
- A chaque itération, on augmente l'angle de rotation de 15 degrés.

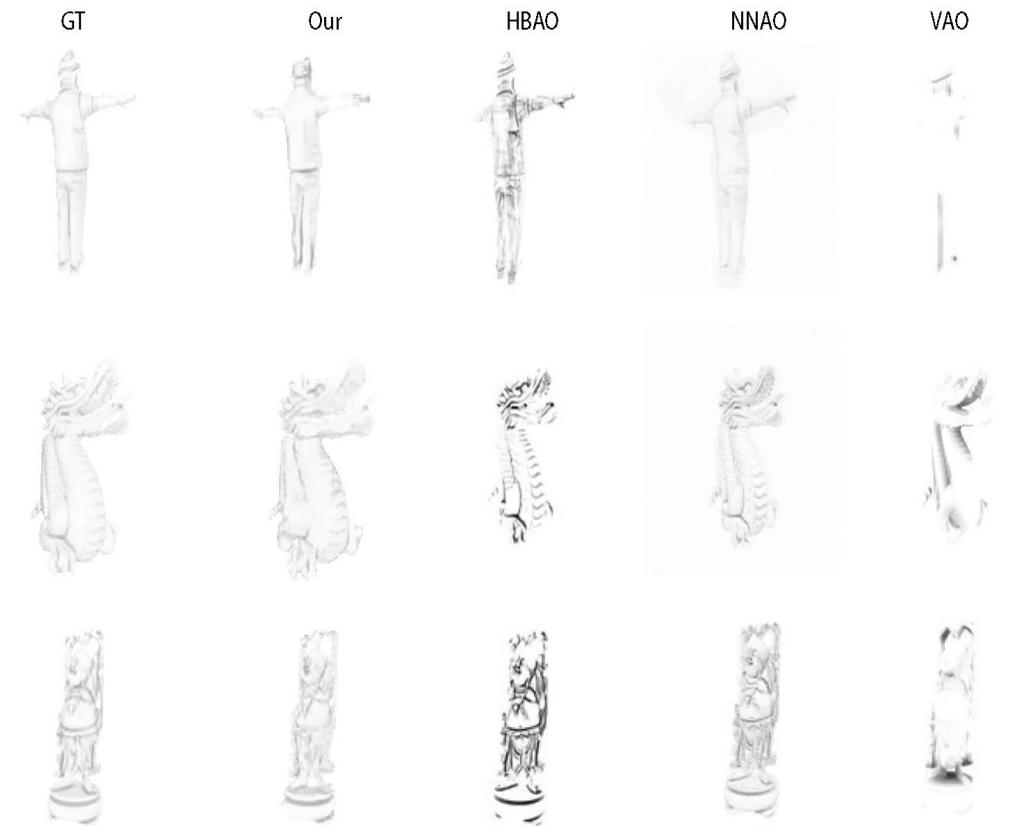
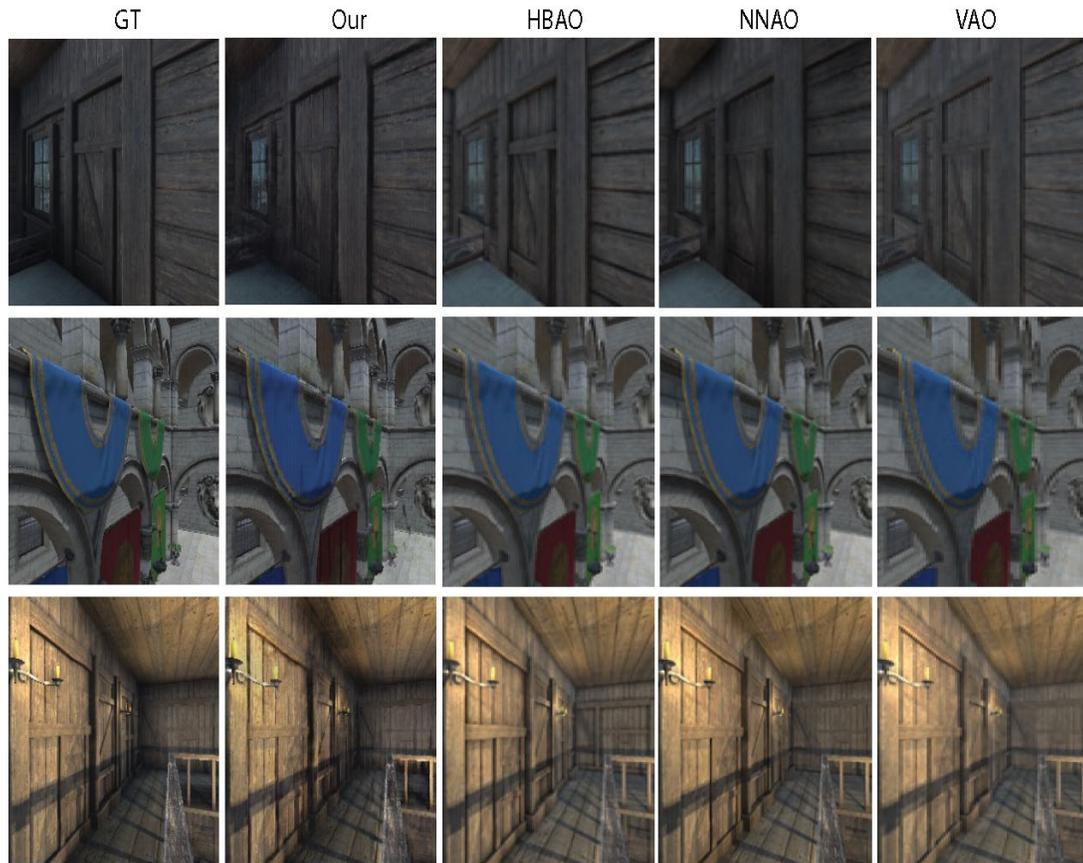


AO without texture



AO with texture

CONTRIBUTION 2

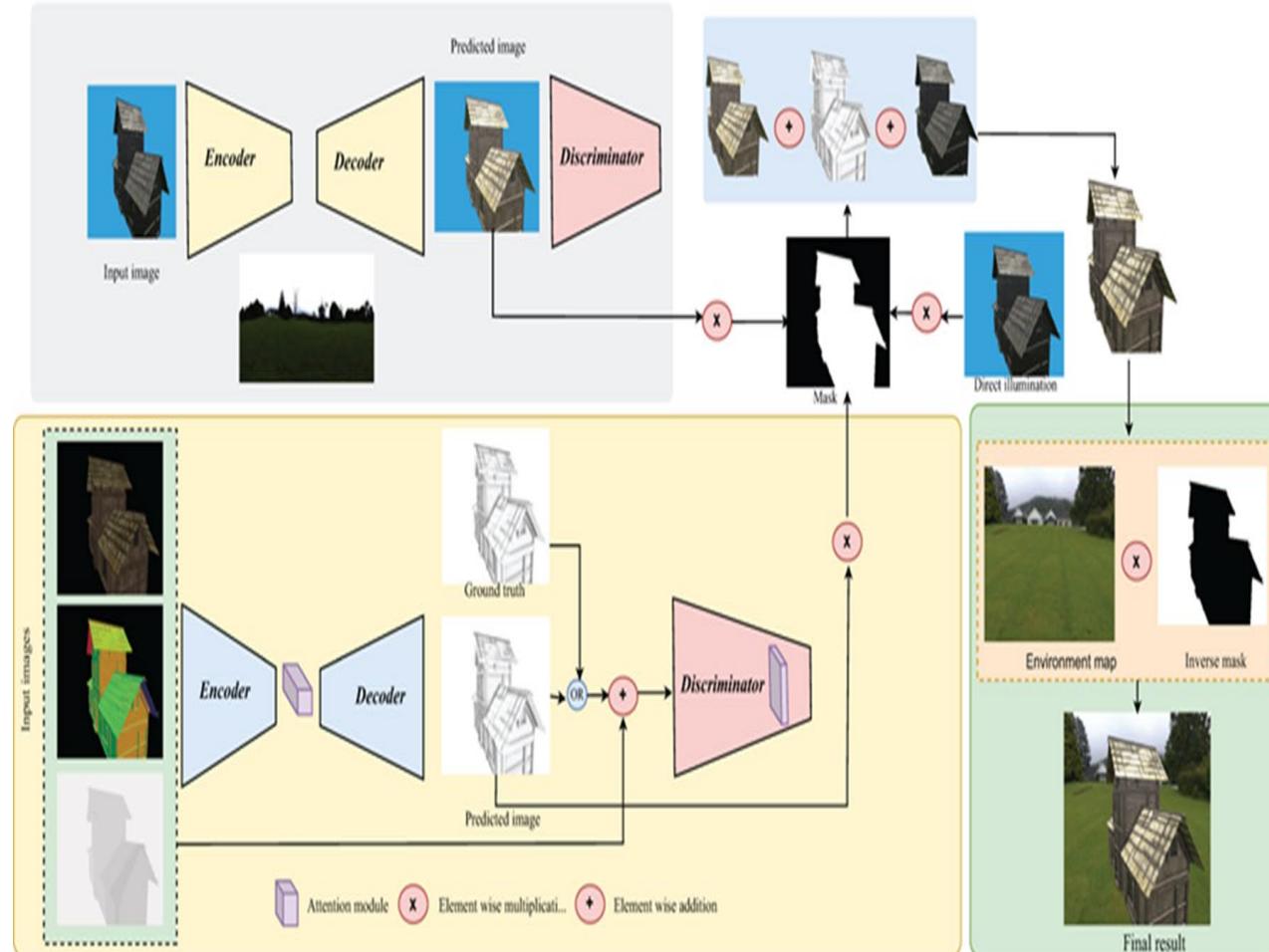


CONTRIBUTION 3



APPROXIMATING GLOBAL ILLUMINATION WITH AMBIENT OCCLUSION AND ENVIRONMENT LIGHT VIA GAN

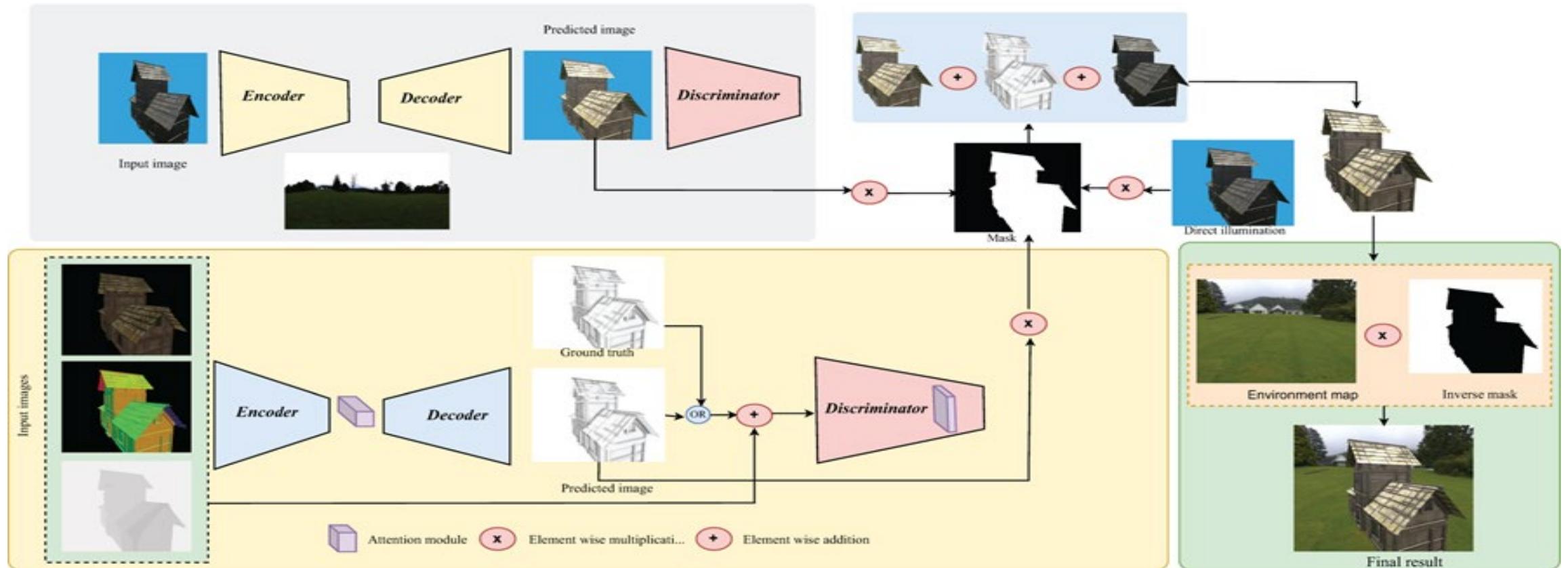
- Approximer l'effet d'illumination globale en utilisant deux réseaux neuronaux.
- **le premier réseau neuronal**, est utilisé pour approximer l'effet d'illumination de l'environnement.
- **Le second réseau neuronal**, un réseau antagoniste génératif est basé sur le modèle utilisé dans la contribution 2.
- Nous avons introduit l'utilisation de deux masques pour préserver la position de l'objet dans l'espace écran.
- nous concaténons les résultats des deux réseaux neuronaux et de l'image d'illumination directe pour fournir l'approximation finale de l'illumination globale.



CONTRIBUTION 3



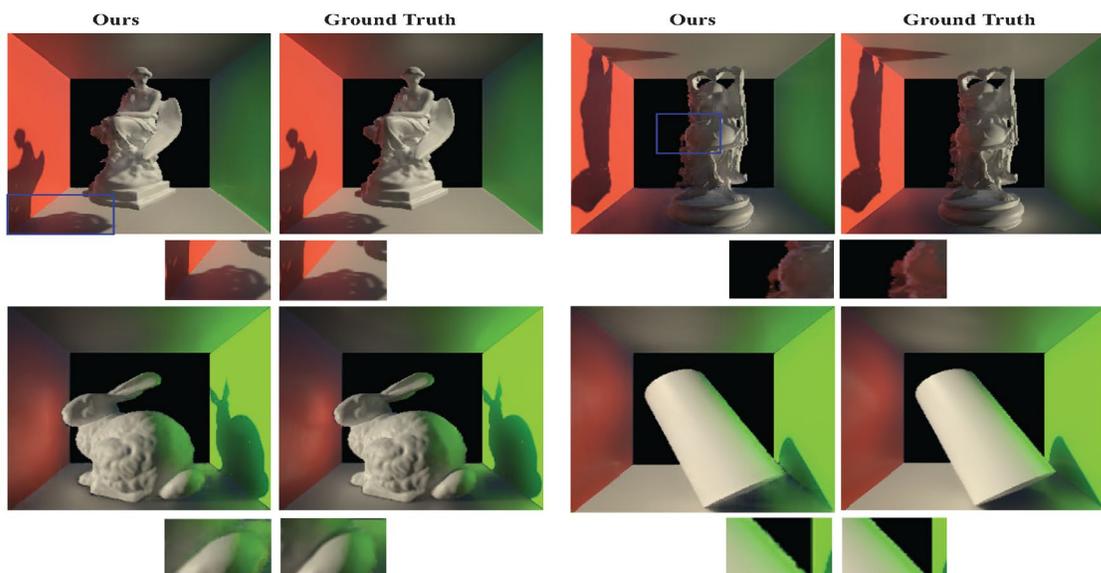
APPROXIMATING GLOBAL ILLUMINATION WITH AMBIENT OCCLUSION AND ENVIRONMENT LIGHT VIA GAN



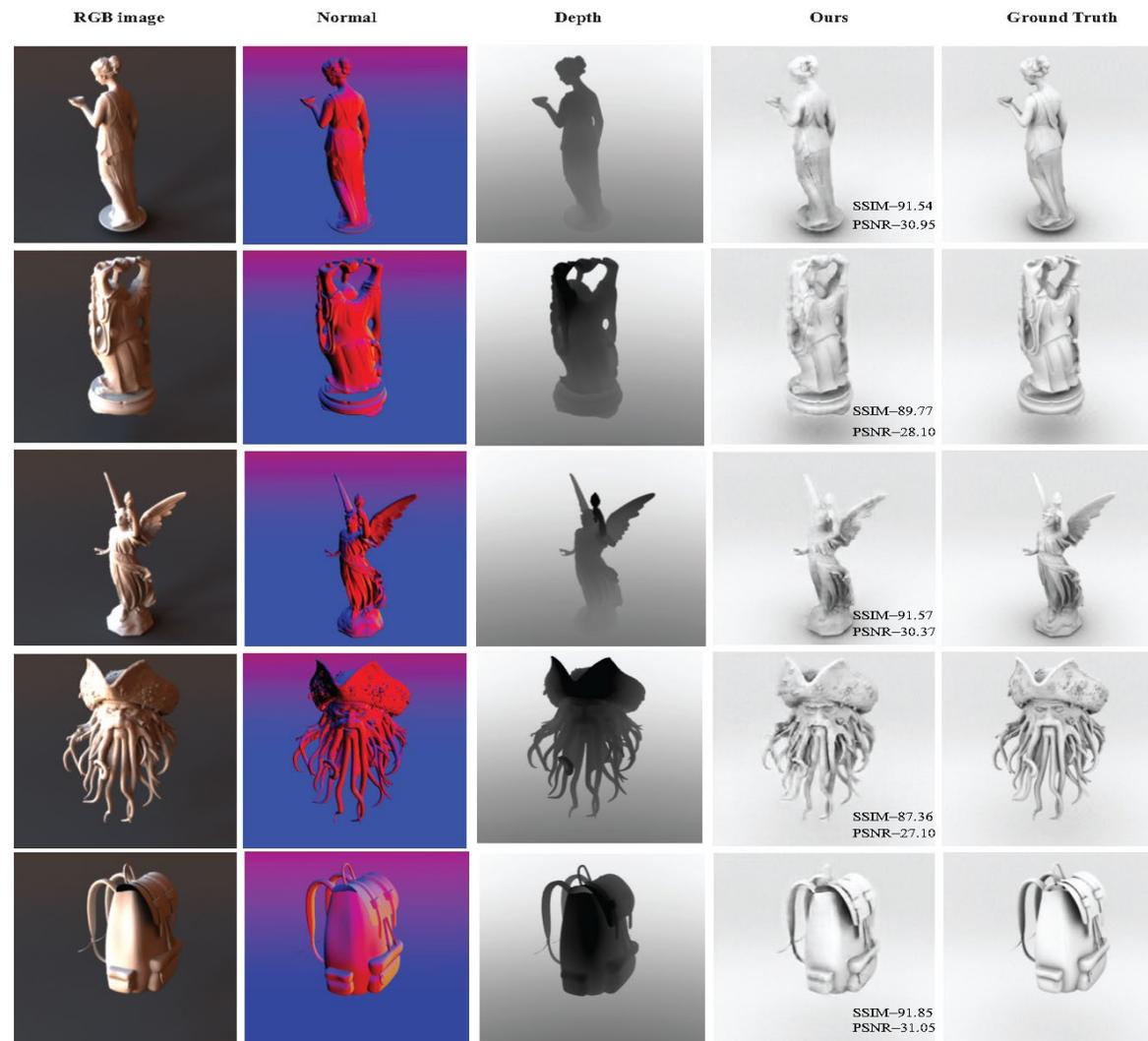
CONTRIBUTION 3



Deep Illumination



Invrrender



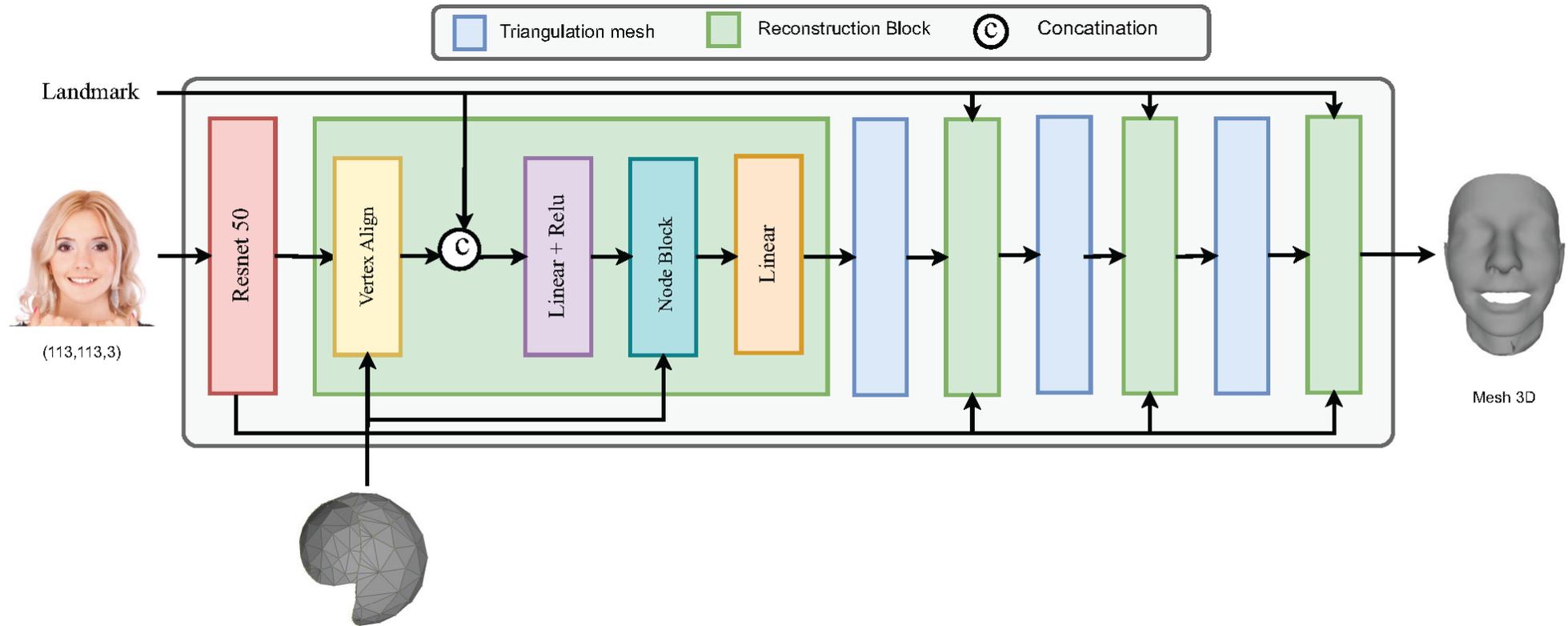
CONTRIBUTION 4



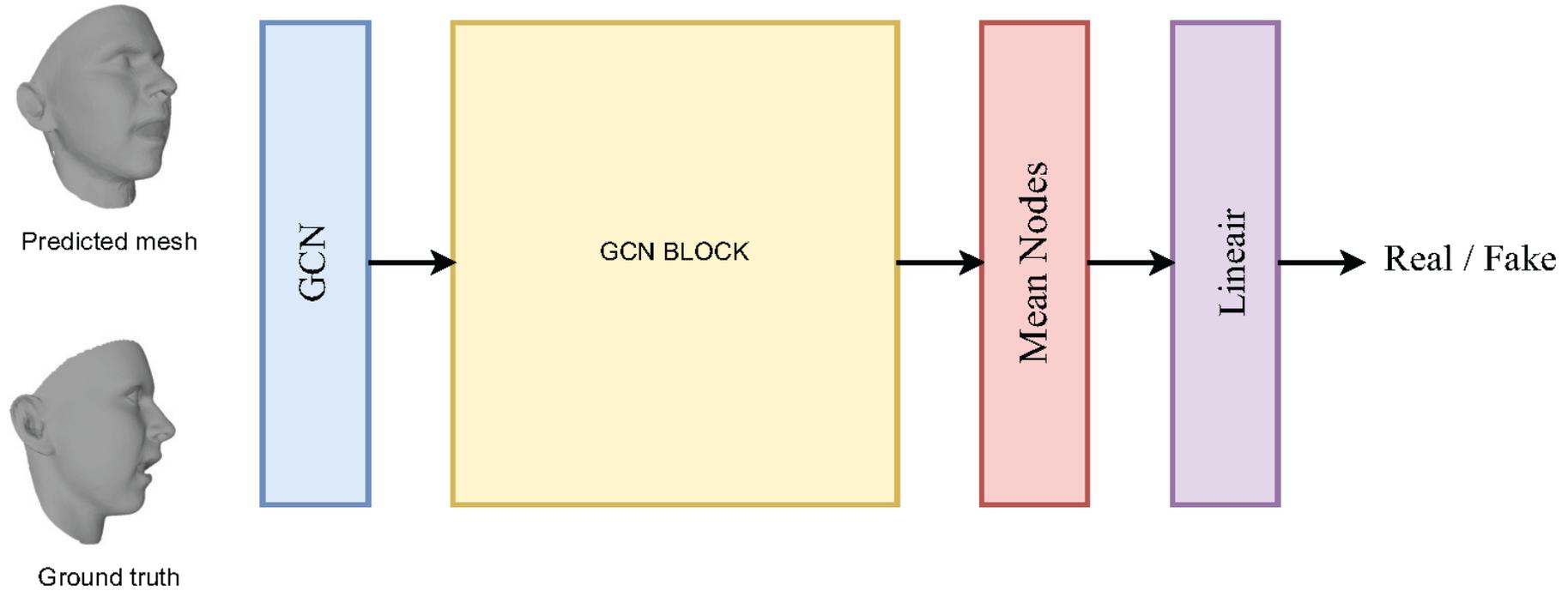
3D FACE RECONSTRUCTION FROM SINGLE IMAGE WITH GENERATIVE ADVERSARIAL NETWORKS

- Reconstruire la géométrie du visage avec expression et position à partir d'une seule Image 2D et points de repère
- Proposer une nouvelle technique basée sur les réseaux de neurones génératifs.
- Introduire une nouvelle architecture pour le discriminateur basée sur des couches de convolution graphique pour augmenter la précision du maillage généré.
- Utilisation des points de repère comme entrée du générateur pour générer un maillage de visage avec expression.

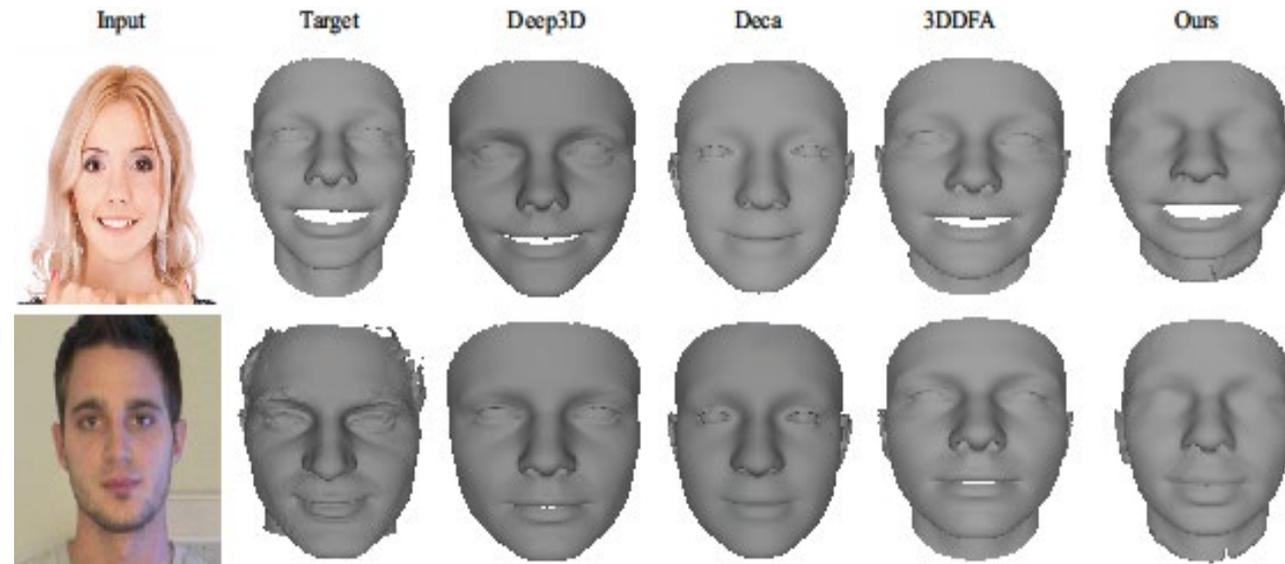
CONTRIBUTION 4



CONTRIBUTION 4



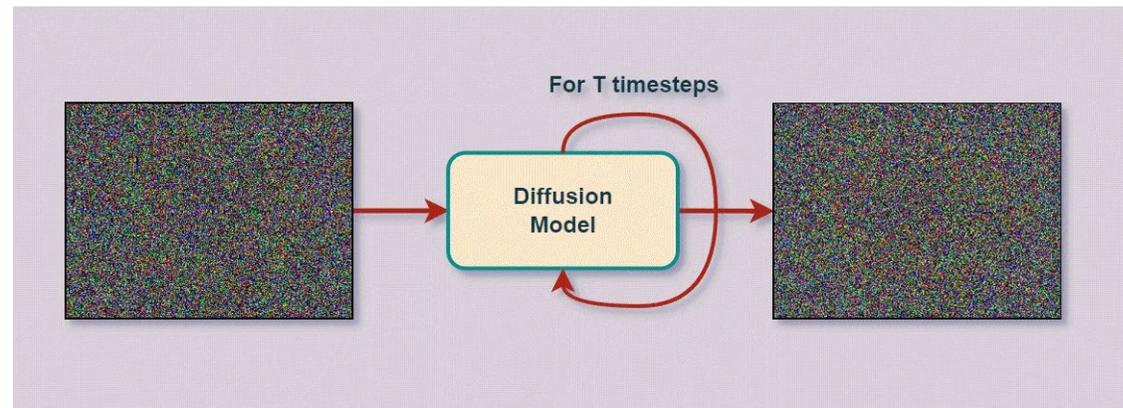
CONTRIBUTION 4



LES MODÈLES DE DIFFUSION

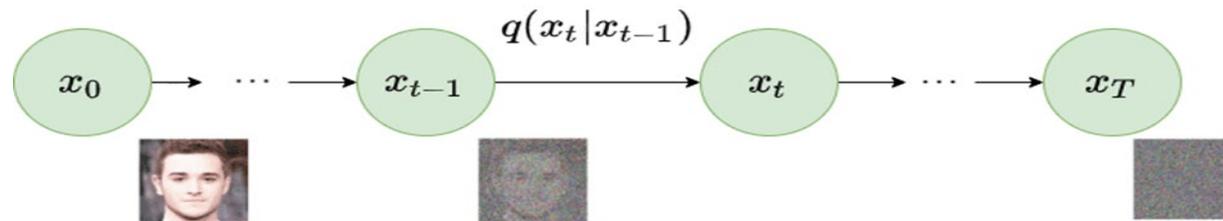


- Les modèles de diffusion sont une classe de modèles génératifs probabilistes qui transforment le bruit en un échantillon de données représentatif.
- Le modèle de diffusion est entraîné pour lier une distribution basique, telle qu'une distribution gaussienne, à une distribution plus complexe.
- Introduisez progressivement du bruit dans l'image initiale.
- Peut générer une variété d'images en modifiant le bruit.



Processus de diffusion vers l'avant

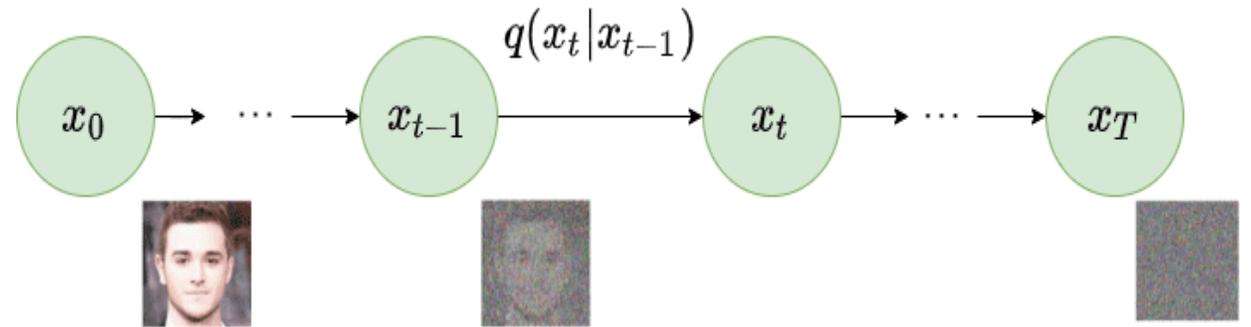
- Le processus de diffusion vers l'avant peut être défini comme une chaîne de Markov où chaque X_t est une version bruitée des données X_0 .
- et le bruit augmente avec le temps. Il peut être représenté comme suit :
$$X_t = \sqrt{1 - \beta_t} \cdot X_{t-1} + \sqrt{\beta_t} \cdot \epsilon_t$$
- où :
- X_t : Les données à l'étape de temps t .
- β_t : Le niveau de bruit à l'étape de temps t .
- ϵ_t : Le bruit ajouté à l'étape de temps t , généralement tiré d'une distribution normale standard.



Processus de diffusion vers l'avant

Le bruit est souvent supposé être gaussien, conduisant à une distribution q

$$q(X_t|X_{t-1}) = \mathcal{N}(X_t; \sqrt{1 - \beta_t} \cdot X_{t-1}, \beta_t \cdot I)$$



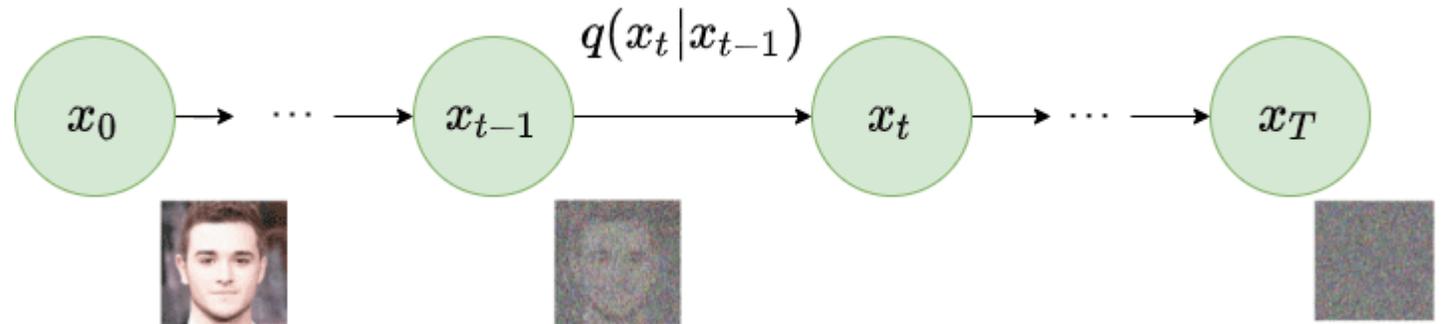
où :

- \mathcal{N} : La distribution normale.
- I : La matrice identité, assurant que le bruit est ajouté indépendamment à chaque dimension des données.

le processus de diffusion inverse

- le processus de diffusion inverse est souvent décrit par une série de distributions conditionnelles.
- Le processus commence avec une image de bruit
- la raffine itérativement pour générer un échantillon qui ressemble aux données d'entraînement.

$$p(\mathbf{x}_0|\mathbf{x}_T) = \prod_{t=T}^1 p(\mathbf{x}_{t-1}|\mathbf{x}_t)$$



APPLICATION TEXT2IMAGE



- **Texte** : un appartement moderne avec de grandes fenêtres à Vienne, couleurs intérieures. vert, crème, blanc et beaucoup de bois



APPLICATION TEXT2IMAGE



- **Texte** : Un appartement africain étrangement intact, une tasse de café à moitié remplie, un journal soigneusement plié sur la table, le tout enveloppé d'un sombre mystère, réaliste.



APPLICATION TEXT2IMAGE



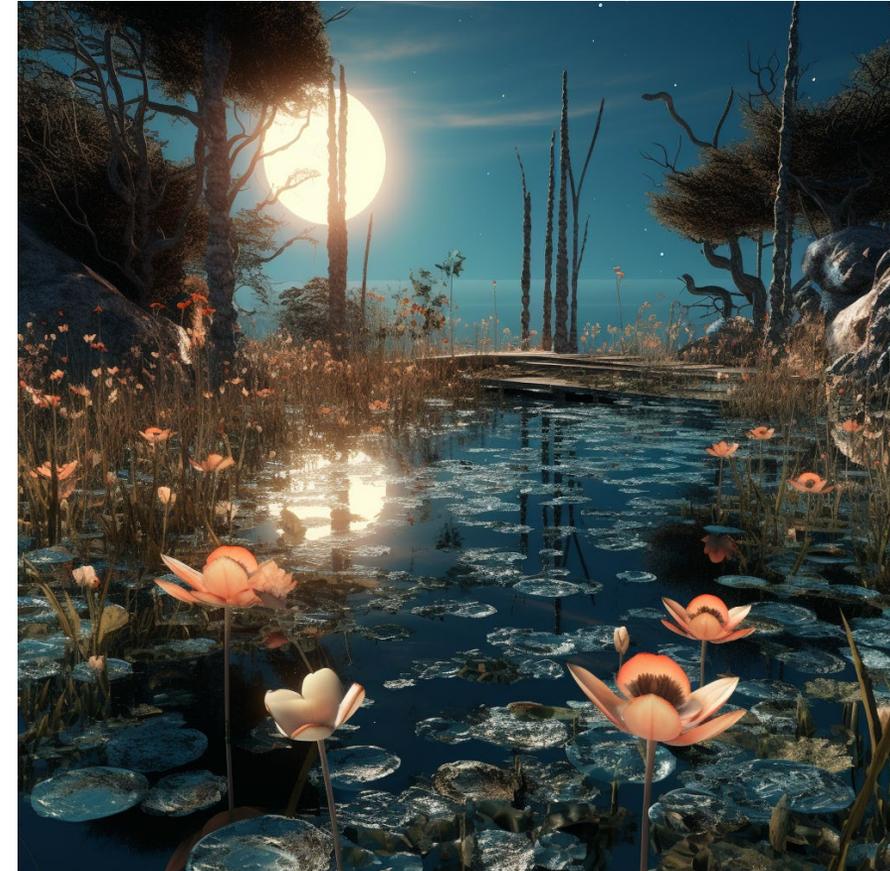
- **Texte** : style dessin animé, rendu pixar, hyper réaliste, 3D



APPLICATION TEXT2IMAGE



- Texte : métaverse du lac des souvenirs



CONCLUSION



- L'efficacité et la qualité des images produites par les modèles génératifs soulignent leur potentiel en tant que solution alternative aux approches conventionnelles.
- Le rendu neuronal offre une alternative prometteuse aux méthodes de rendu traditionnelles.
- Les processus de diffusion démontrent une capacité exceptionnelle à transformer des distributions simples en images complexes et détaillées.



**MERCI POUR VOTRE
ATTENTION**