



Faculté des Sciences de Kénitra
Master Intelligence Artificielle et Réalité Virtuelle

Rapport de Projet

Réalisé Par :

*Walid Abdellah,
Touiyate Adnane
Gargouri Aicha et
Bouktitiya Hamza*

Encadré Par :

Pr. NOURI ANASS

Génération d'un objet 3D à partir d'une image 2D

Année Universitaire : 2023/2024

Table des matières

<u>Table des matières</u>	2
<u>Table des figures</u>	4
<u>Introduction</u>	5
1. Périmètre du projet :	5
1.1 Objectifs :	5
2. Exigences techniques :	5
3. Applications potentielles :	6
4. Etat de l'art :	7
<u>Chapitre 1 : Existant et Architecture du projet</u>	9
1. Les existants :	9
2. Architecture du projet :	10
<u>Chapitre 2 : Génération des Multi-Vues à partir d'une image 2D</u>	12
1. Introduction :	12
2. Modèle Zero-1-to-3 :	12
2.1 Objectif du modèle :	12
2.2 Méthodologie :	13
2.3 Application :	14
3. Conclusion :	15
<u>Chapitre 3 : Génération d'un objet 3D</u>	16
1. Introduction :	16
2. Principes de NeRF :	16
2.1 Représentation volumétrique :	17
2.2 Réseau de neurones :	17
2.3 Entraînement supervisé :	17
2.4 Raymarching :	17
2.5 Effets de vue :	17
3. Principes de COLMAP :	17
3.1 Extraction de caractéristiques avec COLMAP :	18
3.2 Correspondance de caractéristiques avec COLMAP :	18
3.3 Sparse 3D Reconstruction avec COLMAP :	18
3.4 Conversion au format PLY avec COLMAP :	18
3.5 Traitement de nuages de points avec Open3D :	19
3.6 Reconstruction de surface de Poisson :	19

4. Comparaison des résultats :	19
5. Conclusion :	19
<u>Chapitre 4 : Limites du projet</u>	21
1. Exigences élevées en ressources :	21
2. Besoin important en stockage :	21
3. Temps de traitement prolongé :	21
4. Défis dans l'étalonnage multicaméra :	21
5. Incohérences de texture entre les vues :	21
6. Artéfacts d'éclairage et d'ombre :	21
7. Erreurs d'alignement et d'assemblage :	21
8. Perception limitée de la profondeur :	22
9. Limitations de bruit et de résolution :	22
10. Efficacité computationnelle et optimisation :	22
<u>Conclusion</u>	23
<u>Références</u>	24

Table des figures

Figure 1 - Objaverse dataset	6
Figure 2 - Génération Adversarielle Conditionnelle	7
Figure 3 - Structure à Partir du Mouvement.....	8
Figure 4 - Structure de PointNet.....	9
Figure 5 – Reconstruction avec Im2Avatar	10
Figure 6 – Architecture du projet	10
Figure 7 – Zero-1-to-3	13
Figure 8 - Synthèse d'une nouvelle vue	13
Figure 9 - Reconstruction 3D	14
Figure 10 - Image réelle d'un goblet	15
Figure 12 - 3 des vues résultantes.....	15
Figure 13 - Optimisation de NeRF	16
Figure 14 - Le pipeline de SfM de COLMAP	16

Introduction

1. Périmètre du projet :

1.1 Objectifs :

L'objectif principal de ce projet ambitieux est de concevoir et mettre en œuvre un système d'intelligence artificielle (IA) révolutionnaire qui a la capacité unique de transcender les limitations des images 2D pour créer des représentations tridimensionnelles d'une précision exceptionnelle.

En exploitant des photographies en deux dimensions, notre système ouvre de nouvelles perspectives et étend les horizons de la modélisation 3D, de la réalité virtuelle et de la conception. Il s'agit de franchir une étape technologique majeure en offrant aux utilisateurs la capacité de transformer une image 2D en monde 3D captivant, tout en repoussant les frontières de l'innovation dans un large éventail d'industries, de la création artistique à la simulation immersive.

Ce projet vise à révolutionner la manière dont nous percevons et interagissons avec des images, ouvrant la voie à des applications innovantes et à une créativité sans limite.

2. Exigences techniques :

- **Environnement de travail :**

- **Google Colab :** Google Colab ou Colaboratory est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud.
- **Kaggle :** Kaggle est une plateforme web qui héberge la plus grande communauté mondiale de data science et met à sa disposition des outils puissants et des ressources pour contribuer à tous les progrès de data science. Kaggle propose un environnement de Jupyter Notebooks personnalisable et sans configuration. Des GPU ainsi qu'une grande quantité de données et de code publiés par la communauté sont également disponibles gratuitement.

- **L'acquisition de la base de données :** L'utilisation du vaste ensemble de données Objaverse comprenant plus de 800 000 modèles 3D (et en constante augmentation) accompagnés de légendes descriptives, de tags et d'animations. Objaverse représente

une amélioration par rapport aux référentiels 3D actuels en termes d'échelle, de nombre de catégories et de diversité visuelle des instances au sein d'une catégorie.



Figure 1 - Objaverse dataset

- **L'entraînement du modèle (Deep Learning) :** L'exploitation de techniques avancées de deep learning pour entraîner les modèles 3D.
- **Langage de Programmation :** Utilisation de langages de programmation tels que Python.

3. Applications potentielles :

Les applications potentielles de la génération d'objets 3D à partir d'images 2D sont variées et impactent plusieurs domaines. Voici quelques exemples d'applications :

- **Réalité Augmentée (RA) :** La capacité de transformer des images 2D en modèles 3D permet d'améliorer l'expérience utilisateur dans les applications de réalité augmentée. Cela peut inclure la superposition d'objets virtuels dans le monde réel, créant ainsi des expériences interactives et immersives.
- **Jeux Vidéo :** Dans l'industrie du jeu vidéo, la génération automatique de modèles 3D à partir d'images peut accélérer le processus de développement en créant des environnements et des objets de manière plus efficace.
- **Médecine et Chirurgie :** En médecine, la génération d'objets 3D à partir d'images médicales peut aider les chirurgiens à planifier des interventions complexes et à améliorer la visualisation des structures anatomiques.
- **Design Produit :** Les concepteurs de produits peuvent transformer des croquis et des illustrations en prototypes 3D, accélérant ainsi le processus de conception.

Ces applications illustrent la diversité des possibilités offertes par la génération d'objets 3D à partir d'images 2D, ouvrant la voie à de nouvelles innovations et à des expériences utilisateur enrichissantes.

4. Etat de l'art :

L'état de l'art dans le domaine de la génération d'objets 3D à partir d'images 2D est en constante évolution, avec des avancées significatives ces dernières années. Voici quelques approches et techniques importantes utilisées dans cet état de l'art :

- **Génération Adversarielle Conditionnelle (cGAN) :** Les cGAN permettent de conditionner la génération d'objets 3D en fonction d'informations spécifiques, comme des images 2D. Cela permet d'améliorer la précision et le réalisme des modèles générés.

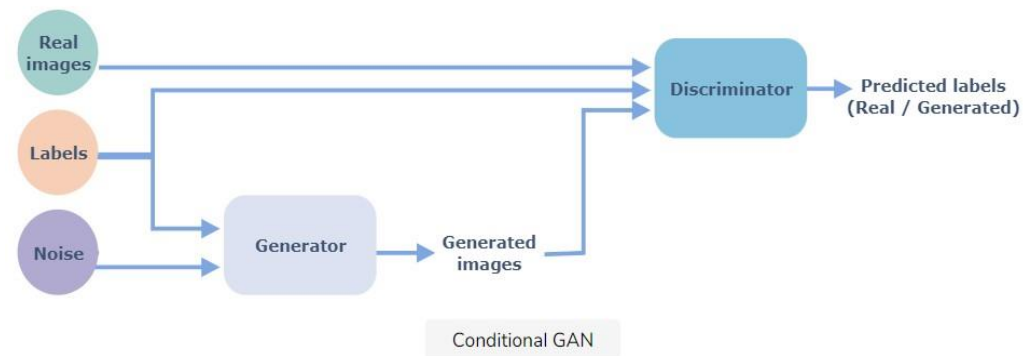


Figure 2 - Génération Adversarielle Conditionnelle

- **Méthodes de Reconstruction Multi-Image :** Certains systèmes exploitent plusieurs images 2D d'un objet sous différents angles pour reconstruire un modèle 3D plus complet. Cela peut être particulièrement utile dans des domaines tels que la vision par ordinateur appliquée à la robotique.
- **Techniques de Structure-à-Partir-du-Mouvement (SfM) :** Ces techniques utilisent des algorithmes pour reconstruire des modèles 3D à partir de séquences d'images 2D prises à différents moments. Elles sont souvent utilisées dans la reconstruction d'environnements et de scènes complexes.

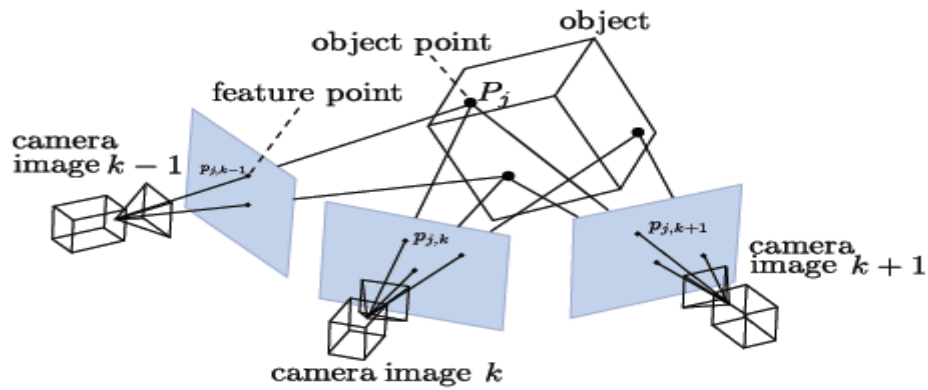


Figure 3 - Structure à Partir du Mouvement

- **Génération de Nuages de Points** : Certains algorithmes se concentrent sur la génération de nuages de points, qui peuvent ensuite être utilisés pour créer des modèles 3D. Ces méthodes sont souvent employées dans la reconstruction de scènes en extérieur.

L'état de l'art reflète une combinaison de méthodes traditionnelles et d'approches basées sur l'apprentissage profond, démontrant la diversité des stratégies pour générer des objets 3D à partir d'images 2D. Les chercheurs continuent d'explorer de nouvelles idées pour améliorer la précision, la robustesse et l'efficacité de ces systèmes.

Chapitre 1 : Existant et Architecture du projet

1. Les existants :

L'existant dans la génération d'objets 3D à partir d'images 2D comprend les différentes approches et méthodes actuellement utilisées dans ce domaine. Quelques exemples d'existants :

- **PIX3D** : Un modèle basé sur un GAN conditionnel qui génère des modèles 3D à partir d'images 2D en prenant en compte des informations supplémentaires telles que la catégorie de l'objet.
- **PointNet** : Bien que principalement utilisé pour la classification d'objets 3D, PointNet peut également être adapté pour la génération d'objets 3D en prenant des images 2D comme entrée et en générant des nuages de points 3D.

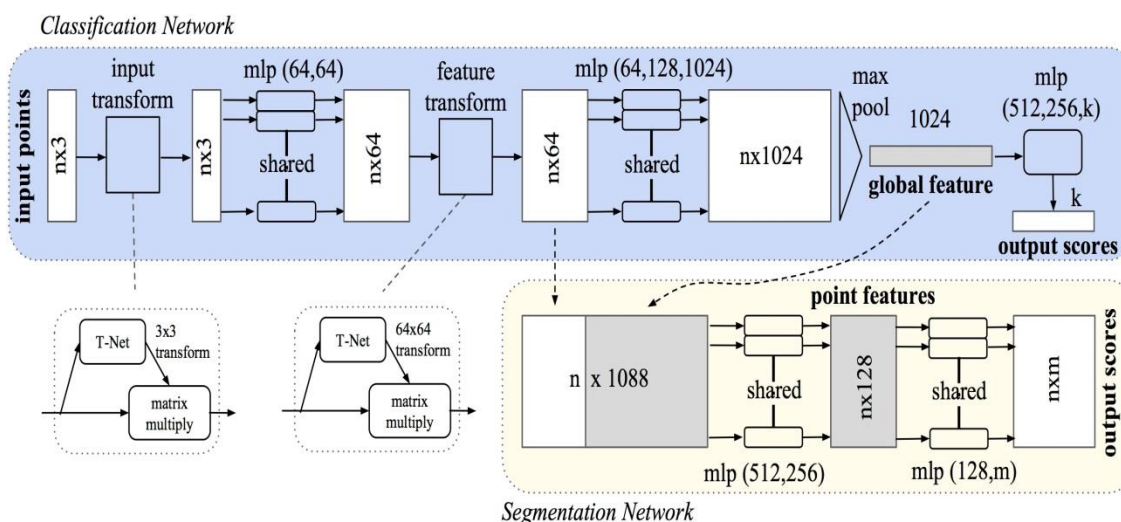


Figure 4 - Structure de PointNet

- **Im2Avatar** : Cette approche utilise une combinaison de CNN et de GAN pour générer des modèles 3D à partir d'images 2D, en mettant l'accent sur la génération d'objets complexes avec des détails fins.

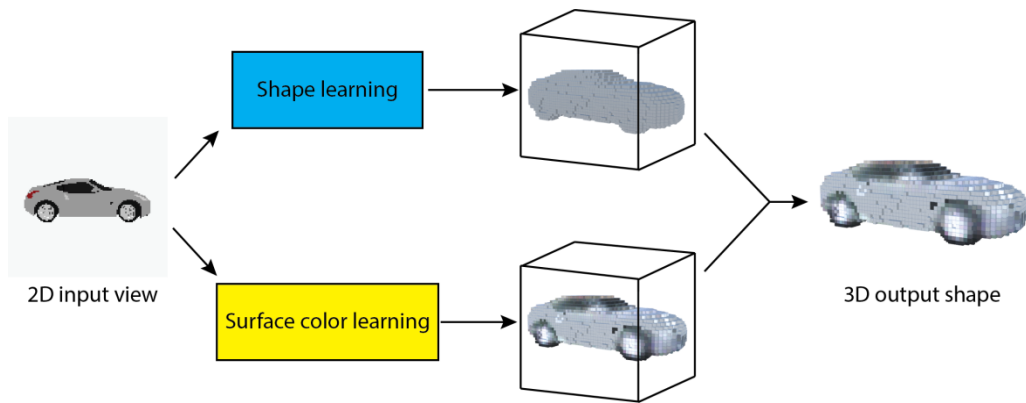


Figure 5 – Reconstruction avec Im2Avatar

L'état actuel de la technologie et les défis rencontrés font partie intégrante de l'existant dans ce domaine, et ils servent de base pour les développements futurs.

2. Architecture du projet :

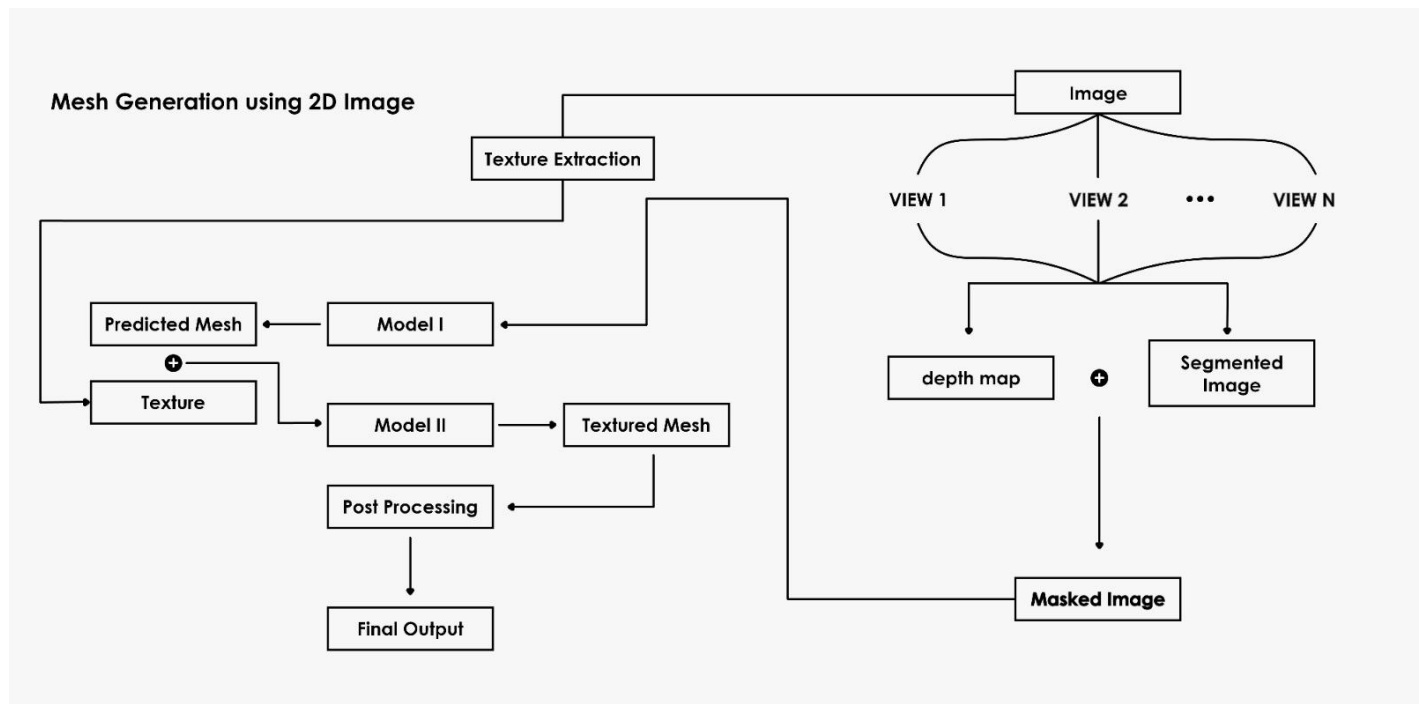


Figure 6 – Architecture du projet

- **Prétraitement des données** : Nous prétraitons l'image pour normaliser la taille, l'orientation et l'éclairage et supprimer l'arrière-plan.
- **Génération de la Vue d'Objet** :
 - **Réseau de Générateur** : Nous créons un réseau neuronal qui génère des vues d'objet. Ce réseau prend une image en entrée et produit une représentation tridimensionnelle de l'objet sous différents angles.

- **Entraînement :** Nous formons le réseau de générateur à l'aide d'un jeu de données d'images 2D et de leurs représentations tridimensionnelles similaires.
- **Résultat :** Le réseau de générateur doit produire plusieurs vues de l'objet.
- **Génération de la Maille et de la Texture de l'Objet :**
 - Une fois que nous avons généré de multiples vues de l'objet, nous pouvons utiliser des techniques telles que la reconstruction stéréo multi-vues (MVS) ou des algorithmes de reconstruction de surface 3D pour créer une représentation tridimensionnelle de maille de l'objet.
 - Nous pouvons également générer des textures pour le modèle 3D en utilisant les informations de couleur des images d'origine.

Chapitre 2 : Génération des Multi-Vues à partir d'une image 2D

1. Introduction :

La génération des multi-vues implique la création de plusieurs perspectives d'un objet à partir d'une seule image bidimensionnelle, offrant ainsi une représentation tridimensionnelle plus complète.

Cette technique est d'une importance capitale dans le domaine de la génération d'objets 3D, car elle contribue à résoudre le défi complexe de créer des modèles tridimensionnels détaillés à partir de données en deux dimensions.

Mais avant de transmettre l'image au modèle zero-1-to-3, nous extrayons d'abord l'objet de l'arrière-plan en utilisant SAM (Segment Anything Model).

2. Modèle Zero-1-to-3 :

2.1 Objectif du modèle :

"Zero-1-to-3" est un framework pour changer le point de vue d'une caméra sur un objet en se basant uniquement sur une seule image RGB. Ce modèle de diffusion conditionnelle utilise un ensemble de données synthétique pour apprendre les commandes du point de vue relatif de la caméra, ce qui permet de générer de nouvelles images du même objet sous une transformation de caméra spécifiée.

Même s'il est entraîné sur un ensemble de données synthétique, le modèle conserve une forte capacité de généralisation zero-shot aux ensembles de données hors distribution ainsi qu'aux images dans la nature, y compris les peintures impressionnistes. Cette approche de diffusion conditionnée par le point de vue peut également être utilisée pour la tâche de reconstruction 3D à partir d'une seule image.

Des expériences qualitatives et quantitatives montrent que cette méthode surpasse significativement les modèles état de l'art de la reconstruction 3D à partir d'une seule vue et de la synthèse de nouvelle vue en tirant parti de l'entraînement préalable à l'échelle d'Internet.

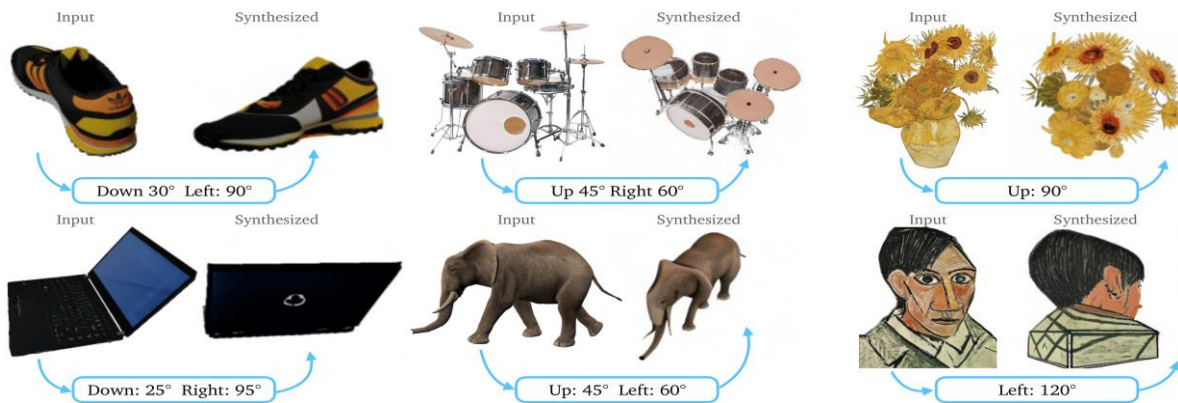


Figure 7 – Zero-1-to-3

2.2 Méthodologie :

La méthodologie du modèle "zero-1-to-3" implique la génération d'images tridimensionnelles (3D) à partir d'images en deux dimensions (2D) à l'aide d'une approche de diffusion stable. Voici les étapes clés de la méthodologie :

- **Synthèse de Nouvelles Vues :** L'objectif principal du modèle est de synthétiser une image d'un objet à partir d'un point de vue différent de la caméra. Cette tâche est formulée comme une synthèse de nouvelle vue, où une image $\hat{x}_{R,T}$ est générée sous une transformation de la caméra spécifiée par la rotation R et la translation T .

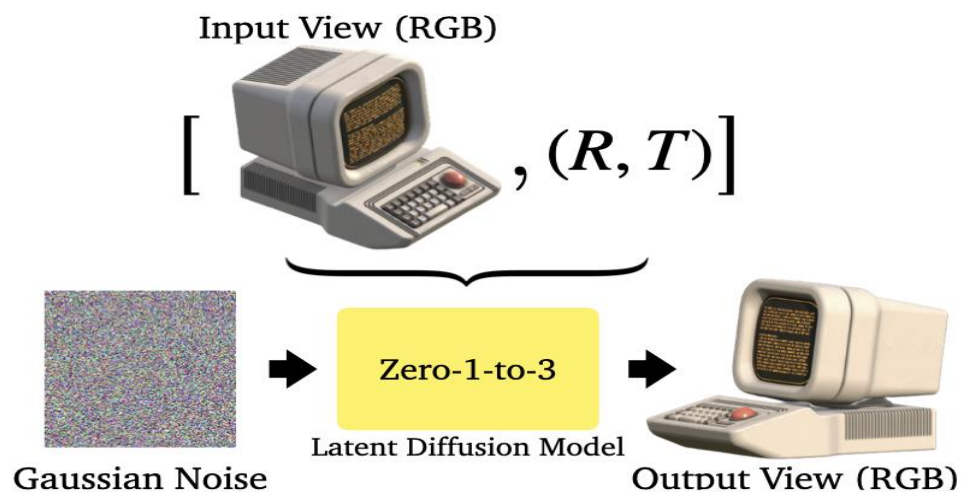


Figure 8 - Synthèse d'une nouvelle vue

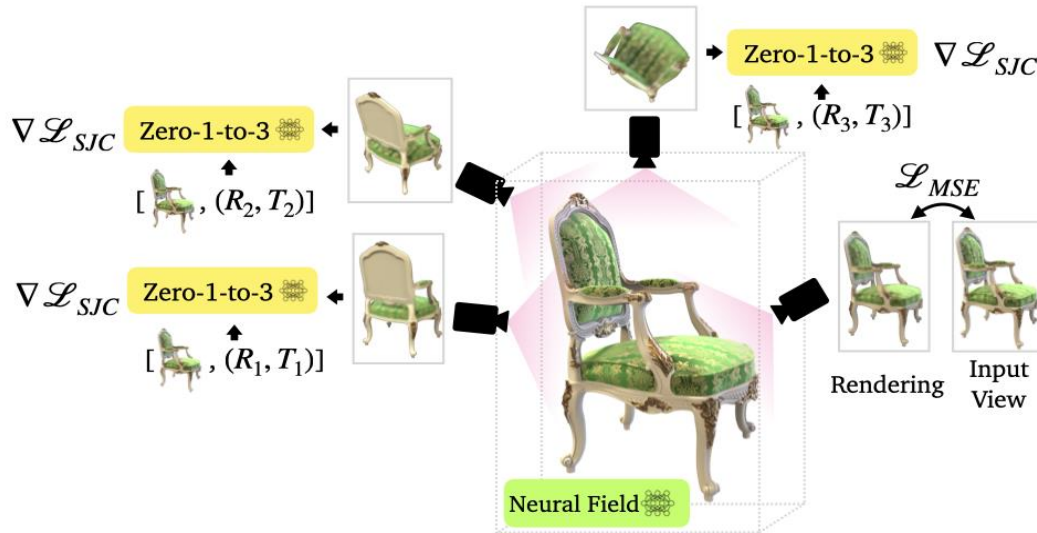


Figure 9 - Reconstruction 3D

- **Pré-Entraînement sur une Large Diffusion de Données :** Les modèles de diffusion sont pré-entraînés sur une vaste quantité de données, ce qui les rend des représentations état de l'art pour la distribution d'images naturelles. Cette étape vise à capturer la variété visuelle des objets et des scènes.
- **Défis à Surmonter :** Deux principaux défis doivent être surmontés. Tout d'abord, bien que les modèles soient formés sur divers objets et points de vue, ils ne capturent pas explicitement les correspondances entre les points de vue. Deuxièmement, les modèles héritent de biais de points de vue provenant des données d'entraînement, entraînant par exemple une tendance à générer des images d'objets orientés vers l'avant.
- **Objectif de Perceptualité :** L'objectif est que l'image synthétisée $\hat{x}_{R,T}$ soit perceptuellement similaire à une vue nouvelle non observée $x_{R,T}$ pour assurer la qualité visuelle et la fidélité à la réalité.

2.3 Application :

- **Input :** Nous donnons en entrée une image 2D réelle comme suit :



Figure 10 - Image réelle d'un goblet

- **Résultat :** Il nous résulte 40 images de différents angles, nous donnons quelques rendements obtenus :



Figure 11 - 3 des vues résultantes

3. Conclusion :

Le modèle Zero-1-to-3 offre une avancée significative dans le domaine de la génération de vues multiples à partir d'une seule image. Grâce à sa capacité à apprendre des contrôles géométriques à partir d'un ensemble de données synthétiques, le modèle démontre une forte généralisation, même lorsqu'il est confronté à des images en conditions réelles.

Les résultats obtenus montrent une nette amélioration par rapport aux modèles existants, tant en termes de qualité que de réalisme des images générées.

Chapitre 3 : Génération d'un objet 3D

1. Introduction :

La génération d'objets 3D à partir de multiples vues offre des avantages significatifs en termes de qualité et de détail, permettant ainsi de capturer fidèlement la géométrie et l'apparence des objets sous différents angles.

Nous allons voir tout d'abord les principes de NeRF qui n'a pas abouti à de bons résultats et c'est pour cela que nous allons utiliser COLMAP dont nous verrons aussi les principes fondamentaux pour ensuite comparer les résultats obtenus de chacun.

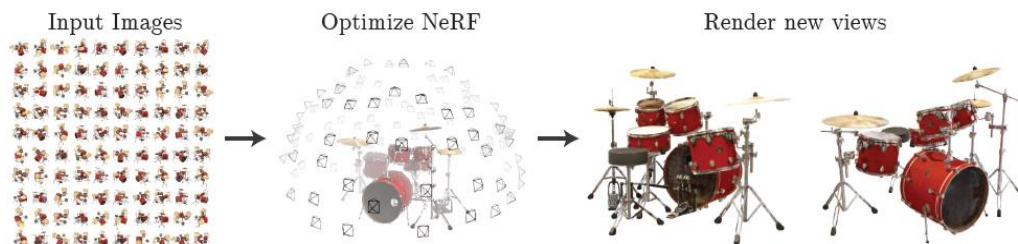


Figure 12 - Optimisation de NeRF

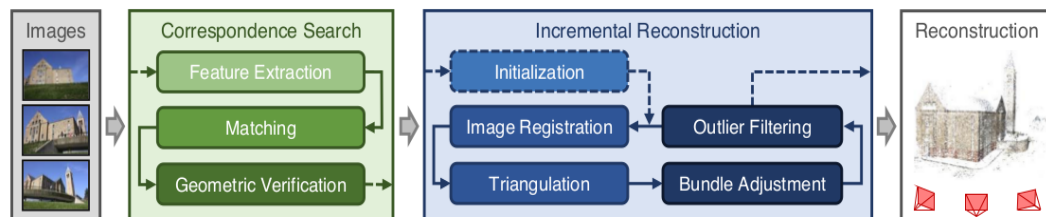


Figure 13 - Le pipeline de SfM de COLMAP

2. Principes de NeRF :

Le modèle Neural Radiance Field (NeRF) est une technique de synthèse d'images 3D qui vise à reconstruire des scènes tridimensionnelles à partir d'un ensemble limité de vues 2D.

Le principe fondamental du modèle NeRF est de représenter une scène en termes de densité et de couleur radiante à chaque point dans l'espace 3D. Contrairement aux approches traditionnelles qui utilisent des maillages ou des nuages de points, NeRF utilise un réseau de neurones pour définir une fonction volumétrique qui mappe les coordonnées 3D vers la densité et la couleur radiante.

Voici les principes clés du modèle NeRF :

2.1 Représentation volumétrique :

La scène est modélisée comme un champ volumétrique continu, où chaque point dans l'espace 3D a une densité et une couleur radiante associées.

2.2 Réseau de neurones :

NeRF utilise un réseau de neurones pour représenter la fonction volumétrique. Ce réseau prend en entrée les coordonnées 3D d'un point et renvoie la densité et la couleur radiante associées à ce point.

2.3 Entraînement supervisé :

Le modèle est entraîné à partir d'un ensemble d'images 2D de différentes vues de la scène. Ces images sont accompagnées de paires de points 3D et de couleurs correspondantes. Le modèle ajuste ses paramètres pour minimiser la différence entre les valeurs prédites et les valeurs réelles.

2.4 Raymarching :

Pour synthétiser une vue à partir de n'importe quel point de vue, NeRF utilise une technique appelée "raymarching". Il trace des rayons à travers le volume de la scène, utilise le réseau de neurones pour évaluer la densité et la couleur le long de ces rayons, puis combine ces informations pour générer l'image synthétique finale.

2.5 Effets de vue :

Le modèle NeRF est capable de capturer des effets visuels complexes tels que les ombres, les reflets et la diffusion de la lumière, grâce à la nature continue de sa représentation volumétrique.

3. Principes de COLMAP :

COLMAP, acronyme de "Structure-from-Motion (SfM) and Multi-View Stereo (MVS) Collection with Efficient Image-Based Localization", est un logiciel open-source utilisé dans le domaine de la vision par ordinateur et de la reconstruction 3D à partir d'images. Il est principalement utilisé pour reconstruire des scènes tridimensionnelles à partir d'une collection d'images bidimensionnelles prises sous différents angles.

En d'autres termes, COLMAP utilise des techniques de "Structure-from-Motion" pour estimer la géométrie 3D d'une scène à partir de multiples images 2D prises à différents points de vue. Il peut également effectuer la reconstruction 3D d'objets à partir de ces images en

utilisant des techniques de "Multi-View Stereo" pour estimer la forme tridimensionnelle à partir des informations de plusieurs vues.

3.1 Extraction de caractéristiques avec COLMAP :

L'extraction de caractéristiques est un processus où des parties distinctes et informatives d'une image (caractéristiques) sont identifiées et décrites. Ces caractéristiques incluent souvent des points, des contours ou des objets uniques et reconnaissables d'une image à une autre.

COLMAP extrait des caractéristiques de chaque image dans l'ensemble de données. L'algorithme (SIFT - Scale-Invariant Feature Transform) identifie des points clés dans les images et les décrit d'une manière invariante à l'échelle, à la rotation et partiellement invariante aux changements d'éclairage.

3.2 Correspondance de caractéristiques avec COLMAP :

La correspondance de caractéristiques consiste à comparer et à aligner les caractéristiques extraites de différentes images pour trouver des correspondances. Ces correspondances sont cruciales pour comprendre comment différentes images sont spatialement liées les unes aux autres.

COLMAP effectue une correspondance exhaustive, ce qui signifie que chaque paire d'images dans l'ensemble de données est comparée pour trouver des caractéristiques correspondantes. Cela fournit les données nécessaires pour comprendre comment les images se chevauchent et comment elles peuvent être combinées pour former une structure 3D cohérente.

3.3 Sparse 3D Reconstruction avec COLMAP :

La reconstruction 3D crée un modèle 3D à partir d'images 2D en triangulant des points caractéristiques. Cela implique d'estimer la position 3D des points en fonction de leur position dans plusieurs images, en utilisant les principes de la géométrie projective.

La fonction de mappage de COLMAP utilise les correspondances de caractéristiques pour trianguler des points dans l'espace 3D, créant un modèle 3D sparse. Ce modèle représente des points clés de la scène sans rendre entièrement les surfaces et les textures.

3.4 Conversion au format PLY avec COLMAP :

Le format PLY (Polygon File Format) est un format de fichier populaire pour stocker des données 3D. Il peut stocker des informations telles que des sommets, des faces et des données de couleur.

Le modèle sparse est converti au format PLY, le rendant compatible avec une large gamme d'outils de visualisation et de traitement 3D.

3.5 Traitement de nuages de points avec Open3D :

Un nuage de points est une collection de points dans l'espace 3D. Le traitement d'un nuage de points implique souvent l'estimation des normales (vecteurs perpendiculaires à la surface à chaque point), qui sont cruciales pour de nombreuses tâches de traitement 3D.

Open3D charge le fichier PLY et estime ces normales. Les orienter vers l'emplacement d'une caméra peut aider dans certains algorithmes qui dépendent de la direction de la surface, comme la reconstruction de surface.

3.6 Reconstruction de surface de Poisson :

La reconstruction de Poisson est une méthode pour créer une surface lisse et continue à partir d'un nuage de points. Elle résout un problème mathématique (équation de Poisson) pour interpoler une surface sur les points.

Open3D applique cette méthode au nuage de points pour créer un maillage détaillé. Le paramètre 'depth' contrôle la résolution et le détail du maillage résultant.

4. Comparaison des résultats :

Les résultats obtenus avec Neural Radiance Field (NeRF) étaient plutôt décevants dans notre projet. Les images générées présentaient des problèmes tels que des lacunes et des déformations, ce qui nous a poussés à rechercher une alternative.

C'est là que COLMAP est intervenu et a vraiment fait la différence. En utilisant des techniques plus traditionnelles pour reconstruire la scène en 3D à partir des caractéristiques des images, COLMAP a produit des résultats bien meilleurs. Les détails étaient plus précis, et la reconstruction de la scène à partir des images était plus robuste grâce aux fonctionnalités telles que la génération de caractéristiques et la triangulation.

5. Conclusion :

Cette comparaison souligne l'importance de choisir la bonne approche en fonction des exigences spécifiques du projet. Alors que NeRF représente une approche innovante basée sur l'apprentissage profond, COLMAP tire parti de méthodes classiques de vision par ordinateur.

Il est crucial de considérer la nature de la scène, la qualité des données disponibles et les exigences de précision lors du choix de la méthode de reconstruction 3D. En fin de compte,

cette expérience met en évidence l'importance de la flexibilité dans l'exploration et l'ajustement d'approches pour obtenir les meilleurs résultats dans des projets complexes de vision par ordinateur.

Chapitre 4 : Limites du projet

1. Exigences élevées en ressources :

Les projets de transformation d'images en 3D nécessitent souvent des ressources informatiques substantielles, notamment des capacités étendues de RAM et de VRAM GPU, pour traiter efficacement des modèles et des textures complexes.

2. Besoin important en stockage :

Ces projets exigent généralement un espace disque significatif pour stocker des images haute résolution, des modèles 3D et des données de traitement intermédiaires.

3. Temps de traitement prolongé :

La conversion d'images en modèles 3D est une tâche intensivement informatique, entraînant des temps de traitement prolongés, en particulier pour des modèles très détaillés.

4. Défis dans l'étalonnage multicaméra :

Calibrer précisément plusieurs caméras pour une reconstruction 3D cohérente et précise peut être complexe, nécessitant un alignement et une synchronisation minutieux.

5. Incohérences de texture entre les vues :

Différentes vues de caméra peuvent capturer des textures variées en raison des conditions d'éclairage ou des angles, provoquant des incohérences dans le modèle 3D final.

6. Artéfacts d'éclairage et d'ombre :

Un éclairage inconstant sur plusieurs images peut entraîner des artéfacts dans le modèle 3D, affectant le réalisme et la précision des textures et des ombres.

7. Erreurs d'alignement et d'assemblage :

Des erreurs dans l'alignement et l'assemblage des images provenant de vues multiples peuvent entraîner des coutures ou des incompatibilités visibles dans le modèle 3D.

8. Perception limitée de la profondeur :

Certains outils peuvent avoir du mal à percevoir avec précision la profondeur, en particulier dans des scènes avec peu de variations de texture ou de contraste.

9. Limitations de bruit et de résolution :

Le bruit et les limites de résolution de l'image peuvent affecter négativement la qualité de la reconstruction 3D, conduisant à des modèles moins détaillés ou inexacts.

10. Efficacité computationnelle et optimisation :

Équilibrer l'efficacité computationnelle avec la qualité de la reconstruction 3D est un défi constant, car des modèles de meilleure qualité nécessitent plus de puissance de traitement et de temps.

Conclusion

En conclusion, ce projet de génération d'un objet 3D à partir d'images 2D en utilisant le modèle zero-1-to-3 pour la génération de multi-vues et COLMAP pour la reconstruction 3D a abouti à des résultats significatifs et prometteurs. L'utilisation du modèle zero-1-to-3 a permis d'obtenir des représentations riches et informatives des scènes à partir de différentes perspectives, fournissant ainsi une base solide pour la reconstruction tridimensionnelle.

L'étape de génération de caractéristiques avec COLMAP a été cruciale pour extraire des informations clés de chaque image et établir des correspondances entre elles. La triangulation des points caractéristiques a ensuite conduit à la création d'un modèle 3D sparse, offrant une représentation spatiale de la scène à partir des images 2D initiales.

Cependant, des défis ont été identifiés, notamment la dépendance aux caractéristiques extraites par le modèle zero-1-to-3 et la sensibilité aux variations d'éclairage et de perspective. La reconstruction 3D peut présenter des lacunes dans les zones où les caractéristiques sont moins distinctes ou absentes.

Il est important de noter que, malgré l'utilisation initiale de Neural Radiance Field, les résultats obtenus n'étaient pas satisfaisants, ce qui a conduit à l'adoption de COLMAP. Cela souligne la nécessité d'explorer et d'ajuster différentes approches pour optimiser les résultats dans des projets complexes de vision par ordinateur.

Bien que des défis subsistent, le projet ouvre des perspectives passionnantes pour l'amélioration future. L'intégration de techniques avancées de traitement d'image et de modélisation 3D pourrait contribuer à surmonter les limitations actuelles. La sensibilisation aux contraintes et aux possibilités du modèle zero-1-to-3 guidera des ajustements futurs pour améliorer la robustesse du système.

En fin de compte, cette approche combinée de génération d'objets 3D à partir d'images 2D offre un aperçu prometteur des possibilités de la vision par ordinateur et de la reconstruction tridimensionnelle. Malgré les défis, ces résultats préliminaires constituent une base solide pour des recherches futures visant à perfectionner et étendre cette méthodologie innovante.

Références

- [Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, Carl Vondrick, "Zero-1-to-3: Zero-shot One Image to 3D Object", 20 Mars 2023](#)
- [© Copyright 2023, Johannes L. Schoenberger](#)
- Cours ~Initiation au traitement et à l'analyse d'images , Pr. Nouri Anass, 2023
- Cours ~Vision Par Ordinateur , Pr. Chahir Youssef
- [Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis", 3 Aout 2020](#)
- <https://pypi.org/project/open3d/>
- <https://pytorch3d.org/>
- [Matt Deitke et al. , "Objaverse-XL: A Universe of 10M+ 3D Objects", 11 Juillet 2023](#)
- [Chunge Bai, Ruijie Fu and Xiang Gao, "Colmap-PCD: An Open-source Tool for Fine Image-to-point cloud Registration", 9 Octobre 2023](#)
- [Johannes L. Schönberger, Enliang Zheng, Marc Pollefeys, Jan-Michael Frahm, "Pixelwise View Selection for Unstructured Multi-View Stereo", 2016](#)
- [Johannes L. Schonberger, Jan-Michael Frahm, "Structure-from-Motion Revisited", 2016](#)
- [Alexander Kirillov et al. , "Segment Anything"](#)