الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي 08 ماي 1945 بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Specialty: **Intelligence Artificielle et Sciences de Données (IASD)**

Theme

3D Objects Generation Using Advanced Deep Learning Techniques

---

Presented by:
Djamel Eddine MESBOUT **Zakaria BABAOUSMAIL**

Defended on: **02 July 2024**
In front of the jury composed of:

Mr. Souleyman CHAIB	President
Ms. Amina TAOULI	Examiner
Mr. Belkacem KHALDI	Supervisor
Mr. Mohammed BEKKOUCHE	Co-Supervisor

Academic Year : 2023/2024

Abstract

This report introduces a novel two-stage approach for generating accurate 3D building models from textual descriptions, addressing a critical challenge in architectural design and game asset creation. In the first stage, we leverage the power of a fine-tuned Stable Diffusion model, a leading text-to-image synthesis approach, to generate high-fidelity images from textual prompts detailing specific building features. Subsequently, these generated images are processed by InstantMesh, a state-of-the-art 3D object reconstruction model, to generate detailed 3D representations.

To enhance the accuracy and quality of the generated models, we meticulously curated a specialized dataset of building images, employing the advanced annotation and summarization capabilities of LLaVa and LLaMa 3. Our experimental results demonstrate the effectiveness of this dual-stage pipeline in accurately translating textual descriptions into precise 3D models.

While the current framework shows promising results, we acknowledge there is room for further refinement and optimization. This work paves the way for streamlining the creation of 3D building models, offering significant benefits to architects in the early stages of conceptual design and providing valuable assets for the gaming industry.

Keywords : 3D, Computer Vision, Natural Language Processing, Text To Image, Tmage To 3D, Text To 3D

Résumés

Ce rapport présente une nouvelle méthode en deux étapes pour créer des modèles 3D réalistes de bâtiments à partir de données textuelles, résolvant ainsi un défi majeur dans les domaines de la conception graphique et de la création de ressources pour les jeux vidéo. Dans un premier temps, nous utilisons la puissance du modèle Stable Diffusion, une technique de génération d'images avancée, pour générer des images fiables à partir de descriptions textuelles de bâtiments spécifiques. Ensuite, ces images sont traitées avec InstantMesh, une méthode de génération 3D d'état de l'art, pour créer des objets 3D détaillés.

Pour améliorer la précision et la qualité des modèles générés, nous avons soigneusement constitué un ensemble de données d'images et d'objets 3D en tirant parti des capacités avancées d'annotation et de résumé de LLaVa et LLaMa 3. Les résultats de la recherche démontrent l'efficacité de ce réseau pour générer des informations détaillées sur chaque bâtiment.

Bien que le cadre actuel donne des résultats prometteurs, nous reconnaissons qu'il existe des possibilités de raffinement et d'optimisation supplémentaires. Ce travail fournit un moyen de simplifier la création de modèles 3D de bâtiments, ce qui bénéficiera grandement aux architectes dès les premières étapes de la conception et fournira des outils précieux pour la production de jeux vidéo.

Mots-clés : 3D, Vision Par Ordinateur, Traitement Du Langage Naturel, Texte vers Image, Image vers 3D, Texte vers 3D

ملخص

نقدم في هذا التقرير طريقة مبنية على الذكاء الاصطناعي لإنشاء نماذج ثلاثية الأبعاد للمباني من نص وصفي كمدخل. يهدف هذا العمل إلى تسهيل التصميم المعماري للمباني في عدة مجالات، مثل تصميم وإنشاء الألعاب الفيديو. عملية إنشاء النموذج الثلاثي الأبعاد مقسمة إلى مرحلتين: تبدأ المرحلة الأولى بأخذ نص وصفي دقيق للمبنى المرجو إنشاؤه وإدخاله في نموذج ذكاء اصطناعي لإنشاء صورة للمبنى. ثم يتم استخدام تلك الصورة كمدخل لنموذج ذكاء اصطناعي آخر مسؤول عن إنشاء النموذج الثلاثي الأبعاد النهائي.

خلال العمل على هذا المشروع، مررنا بمرحلة إنشاء قاعدة بيانات لاستخدامها في تدريب نماذج الذكاء الاصطناعي. تمثلت هذه المرحلة في جمع صور ونماذج ثلاثية الأبعاد لأنواع مختلفة من المباني، وباستخدام عملية تلقائية، تم إنشاء وصف دقيق لكل مبنى.

على الرغم من أن العمل الحالي يظهر نتائج واعدة، إلا أننا ندرك أن هناك فرصًا لمزيد من التحسينات لهذا العمل الذي سيفيد المهندسين المعماريين بشكل كبير، خاصة في المراحل الأولى من التصميم ، ويوفر أداة قيمة في مجال إنشاء ألعاب الفيديو.

الكلمات المفتاحية

نموذج ثلاثي الأبعاد، رؤية الحاسوب، معالجة اللغة الطبيعية، نص إلى صورة، صورة إلى نموذج ثلاثي الأبعاد، نص إلى نموذج ثلاثي الأبعاد