

EPFL **ETH** zürich

PRS

MASTER THESIS

A Modular Framework for Single-View 3D Reconstruction of Indoor Environments

Author
Yuxiao Li

Advisors
Prof. Frédéric Kaplan
Digital Humanities Lab
Prof. Konrad Schindler
Photogrammetry and Remote
Sensing Lab

Supervisors
Dr. Anton Obukhov
Dr. Xi Wang

Abstract

We propose a modular framework for single-view indoor scene 3D reconstruction, where several core modules are powered by diffusion techniques. Traditional approaches for this task often struggle with the complex instance shapes and occlusions inherent in indoor environments. They frequently overshoot by attempting to predict 3D shapes directly from incomplete 2D images, which results in limited reconstruction quality. We aim to overcome this limitation by splitting the process into two steps: first, we employ diffusion-based techniques to predict the complete views of the room background and occluded indoor instances, then transform them into 3D. Our modular framework makes contributions to this field through the following components: an amodal completion module for restoring the full view of occluded instances, an inpainting model specifically trained to predict room layouts, a hybrid depth estimation technique that balances overall geometric accuracy with fine detail expressiveness, and a view-space alignment method that exploits both 2D and 3D cues to ensure precise placement of instances within the scene. This approach effectively reconstructs both foreground instances and the room background from a single image. Extensive experiments on the 3D-Front dataset demonstrate that our method outperforms current state-of-the-art (SOTA) approaches in terms of both visual quality and reconstruction accuracy. The framework holds promising potential for applications in interior design, real estate, and augmented reality.

Key Words: Single-View Reconstruction, Diffusion Techniques, Amodal Completion, Inpainting Model, Computer Vision

Résumé

Nous proposons une approche modulaire basée sur des techniques de diffusion pour la reconstruction de scènes d'intérieur en 3 dimensions à partir d'une seule image. Les approches traditionnelles rencontrent souvent des difficultés quand elles font face à des objets aux formes irrégulières ou aux occultations inhérentes aux environnements intérieurs. Celles-ci tentent d'approximer directement la forme des objets à partir d'images incomplètes, conduisant à des reconstructions inexactes. Notre approche vise à surmonter cette limitation en divisant le processus de reconstruction en deux étapes: d'abord, nous exploitons des techniques de diffusion afin de prédire avec précision l'arrière-plan de la pièce et les objets d'intérieur occultés. Ces derniers sont ensuite modélisés en 3 dimensions. Nos contributions sont les suivantes: un module de complétion amodale restaurant complètement les instances occultées, un modèle d'inpainting spécifiquement entraîné pour prédire la disposition des pièces, une technique hybride d'estimation de profondeur équilibrant la géométrie globale de l'environnement et les détails de ses éléments, et une méthode d'alignement dans l'espace exploitant à la fois les informations en deux et trois dimensions pour placer avec précision les différents objets de l'environnement. Cette approche permet de reconstruire efficacement à la fois les objets en premier plan ainsi que l'arrière-plans de la pièce à partir d'une seule image. Des expériences sur le jeu de données 3D-Front démontrent que notre approche dépasse l'état de l'art en termes de qualité visuelle et de précision de reconstruction. Cette approche semble prometteuse dans les domaines du design d'intérieur, de l'immobilier et de la réalité augmentée.

Mots clés: Reconstruction à partir d'une seule image, Techniques de diffusion, Complétion amodale, Modèle d'inpainting, Vision par ordinateur