

AI-based Large Scale 3D Terrain Completion

2023 - Current

Project Leader: Dr. Yajie Zhao; GRAs: Haiwei Chen, Wenbin Teng, Gonglin Chen

Input	Ours	Input	Ours ₁	Ours ₂	Ours ₃	Ours ₄

Background

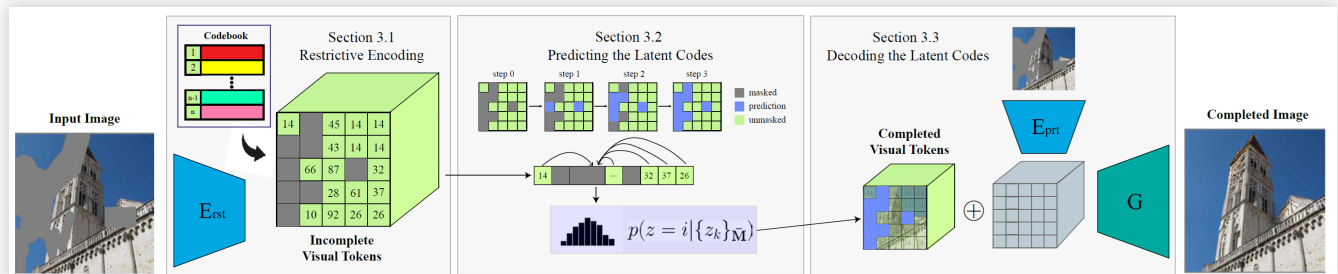
Our method extends from a family of methods characterized by learning priors from discrete latent codes that are obtained from a vector-quantized autoencoder. Past research in this direction has only focused on image synthesis, from directly predicting pixels as word tokens, to predicting tokens encoding visual features of larger receptive fields. While the pioneering works infer latent codes auto-regressively, MaskGIT finds it beneficial to synthesize an image in a scattered manner with a bidirectional transformer. In every iteration, several new codes are predicted in parallel and inserted into scattered locations of the code map until the entire grid is filled. While it has partially adapted its bidirectional framework to the image inpainting setting, our method design addresses several unanswered aspects of this adaptation: how partial images can be robustly masked into latent codes, and how the latent codes should be decoded into synthesized pixels that respect the observable area.

Objectives

Traditional large-scale terrain 3D reconstruction pipelines often result in artifacts and holes in the processed textures and geometries. This typically occurs due to missing views or occlusions, which fail to produce point clouds during the Structure from Motion (SfM) pipeline. Highly skilled artists are then required to manually define and fix these regions before converting the 3D reconstructed terrain into a high-quality simulation environment for soldier training or visualization in path planning. We propose using AI-based methods to post-process raw 3D terrain models obtained from commercial structure-from-motion pipelines. Our aim is to automatically detect, fix, and complete large 3D models in both texture and geometry. This approach will significantly improve efficiency, reduce costs, and simplify the process of providing high-quality 3D environments for downstream use.

Results

VGL achieved the goal by proposing a novel 2D inpainting neural network and integrating a metric depth estimation algorithm into the inpainting process. We begin with automatic hole detection in 2D and apply inpainting trained on one million indoor and outdoor images. Next, we predict monocular depth based on the inpainted 2D image. Finally, we project the single-view depth onto the real-world metric from the SfM pipeline. Using the inpainted depth and texture, our tool produces a completed mesh and texture that integrates smoothly with the original 3D reconstructed models. The proposed inpainting method was accepted by CVPR 2024 as “Don’t Look into the Dark: Latent Codes for Pluralistic Image Inpainting,” which achieved the best inpainting results to date.



Next Steps

VGL plans to propose a more elegant solution for resolving sparse views and occlusions in input images. Instead of running a Structure from Motion (SfM) pipeline to obtain a model and then fill holes in post-processing, we aim to directly generate a complete model. By introducing a generative model into the reconstruction process, we’re combining diffusion models with 3D reconstruction using 3D Gaussian Splatting. This approach will naturally resolve the problem in an end-to-end manner.

Find the referenced publication, “Don’t Look into the Dark: Latent Codes for Pluralistic Image Inpainting” and other publications from the Vision and Graphics Lab at: <https://vgl.ict.usc.edu/publications.php>

Project Leader: Dr. Yajie Zhao; GRAs: Haiwei Chen, Wenbin Teng, Gonglin Chen

Established in 1999, the USC Institute for Creative Technologies (ICT) is a Department of Defense (DoD) University Affiliated Research Center (UARC), sponsored by the US Army. Harnessing Hollywood-derived creativity with academic innovation and military-domain expertise, ICT conducts award-winning R&D in Artificial Intelligence (AI), Computer Graphics, Geospatial Sciences, Human Performance, Learning Sciences, Modeling, Simulation & Gaming, Mixed Reality (MxR), Medical VR, Narrative, and Virtual Humans.