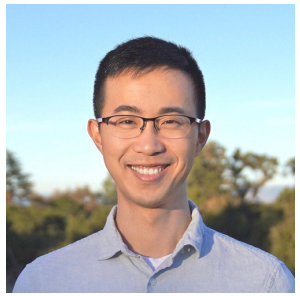
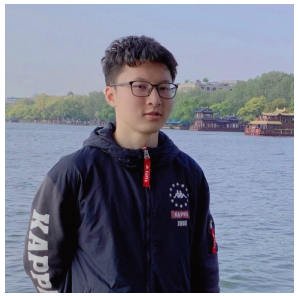
WonderJourney:

Going from Anywhere to Everywhere

Hong-Xing “Koven” Yu Haoyi Duan Junhwa Hur Kyle Sargent Michael Rubinstein

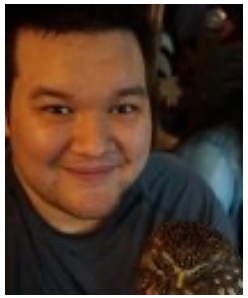
William T. Freeman Forrester Cole Deqing Sun Noah Snavely Jiajun Wu Charles Herrmann







Stanford University Google Research

Motivation: Can we visually imagine Alice’s journey in wonderland?

Motivation: Can we visually imagine Alice's journey in wonderland?

Input: a single image

OR

text ("Girl in wonderland")

Generated "wonderjourney"

Motivation: Can we visually imagine Alice’s journey in wonderland?

Input: a single image

OR

text (“Kids on a farm”)

Generated “wonderjourney”

Problem formulation: Perpetual 3D Scene Generation

Goal: Creating a sequence of diverse yet naturally connected 3D scenes.

Challenges: Generating diverse and plausible scene elements

- Prior works on perpetual view generation only focuses on a single type of scenes.

A. Liu et.al., *Infinite Nature: Perpetual View Generation of Natural Scenes from a Single Image*, 2021

Challenges: Generating diverse and plausible scene elements

- Prior works on perpetual view generation only focuses on a single type of scenes.

- We start from any user-provided location (anywhere), and end at any plausible locations (everywhere).

Serene lake ... →

→

Challenges: Generating diverse and plausible scene elements

- The challenge is to generate diverse and plausible objects, backgrounds, and layouts, that fit into observed scenes and transit to next scene.

- This requires **semantic understanding** (e.g., lion in a kitchen), **visual common sense** (e.g., lion flying in the air), and **geometric understanding** (e.g., disocclusion, parallax, spatial layouts).

Serene lake ... →

→

WonderJourney: A modularized framework

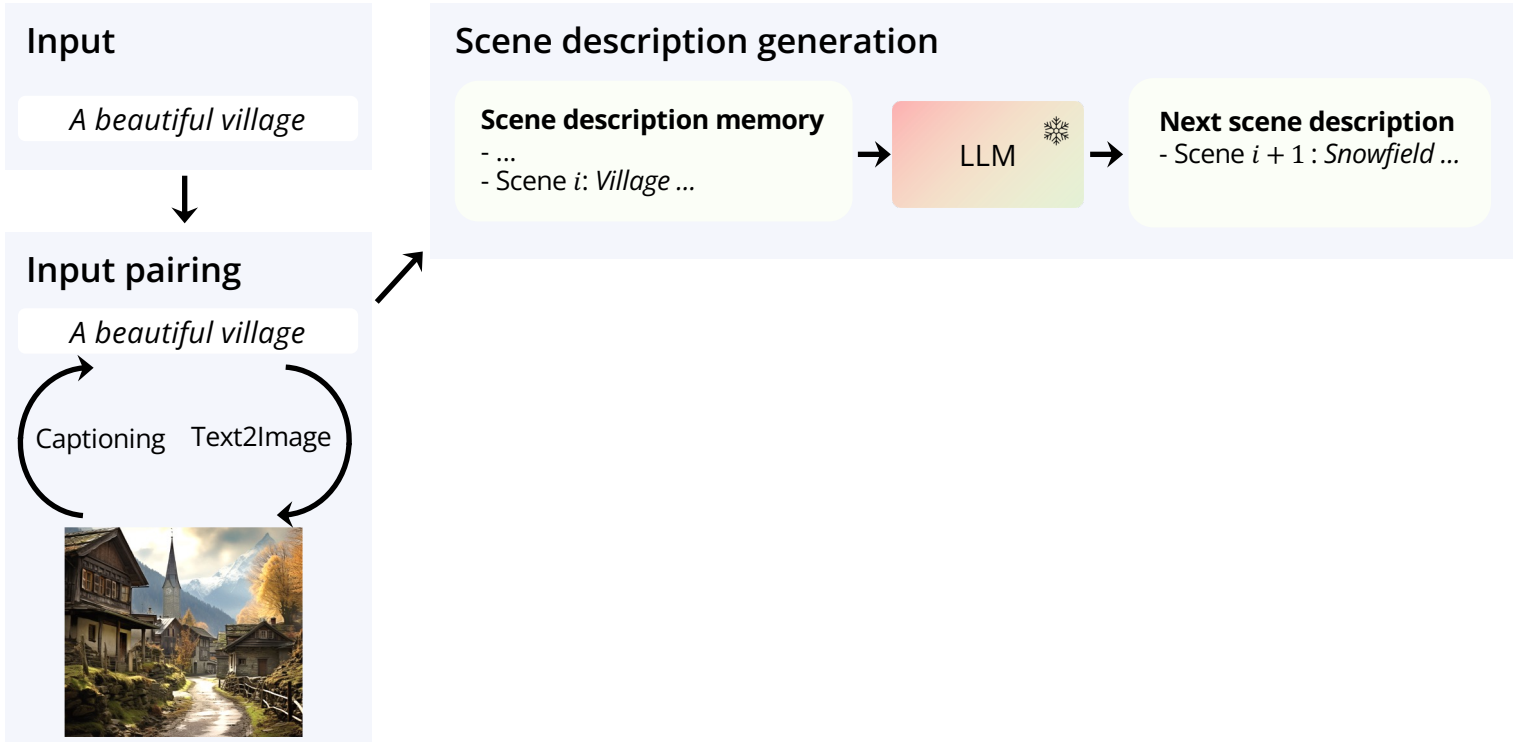
- **Semantic understanding**: Large language model (LLM)

- **Visual common sense**: Text-guided image generator, visual language model (VLM)

- **Geometric understanding**: Depth estimation pipeline, 3D rendering, text-guided image generator

WonderJourney: A modularized framework

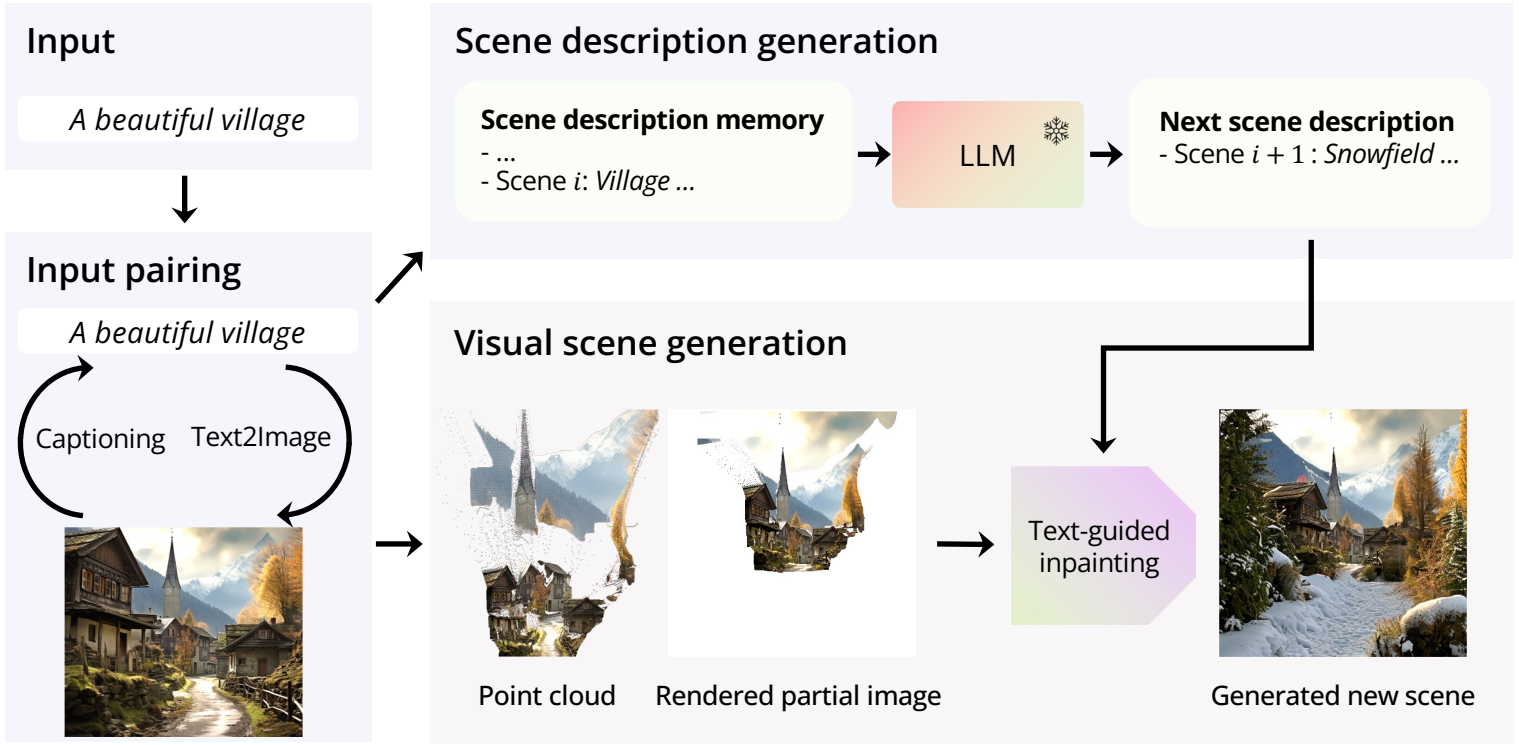
- We use a Large language model (LLM) to generate a long sequence of scene descriptions.



WonderJourney: A modularized framework

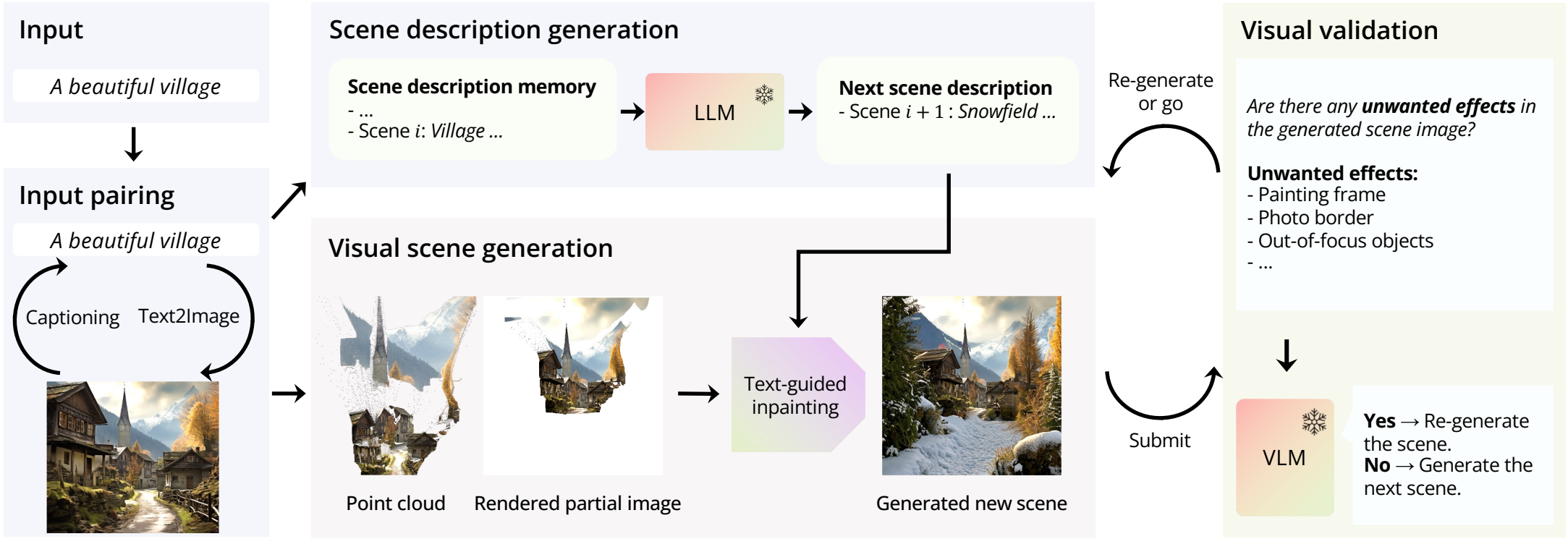
- We use a Large language model (LLM) to generate a long sequence of scene descriptions.

- A text-driven point cloud generation pipeline to synthesize 3D visual scenes.



WonderJourney: A modularized framework

- We use a Large language model (LLM) to generate a long sequence of scene descriptions.
- A text-driven point cloud generation pipeline to synthesize 3D visual scenes.
- A large Vision-Language model (VLM) to verify the generated scenes.

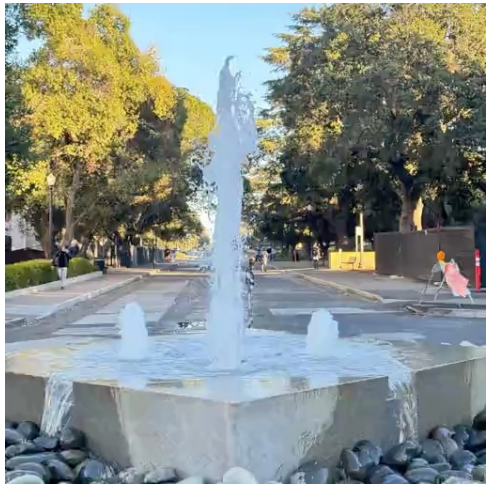


Results: From anywhere

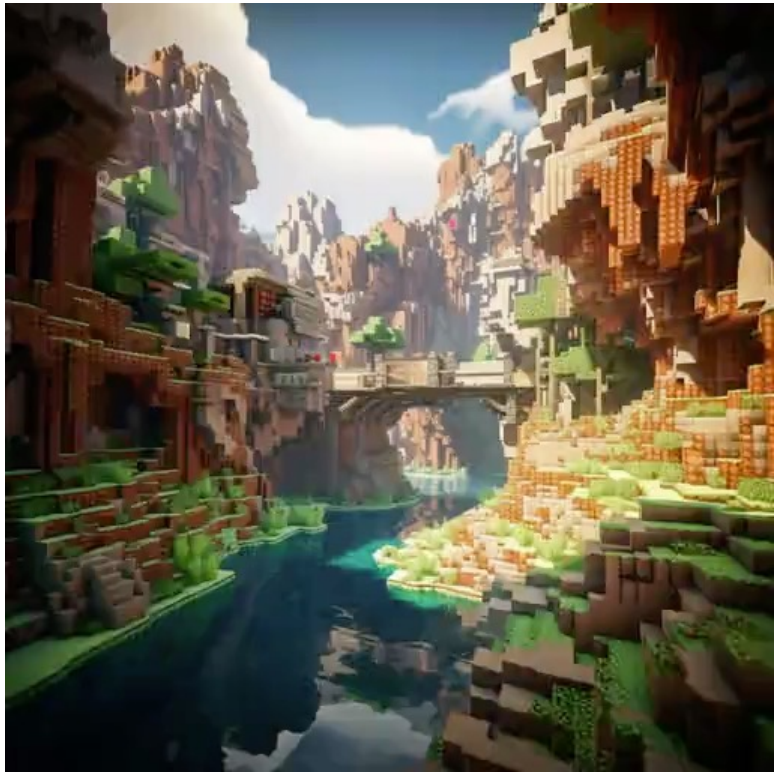
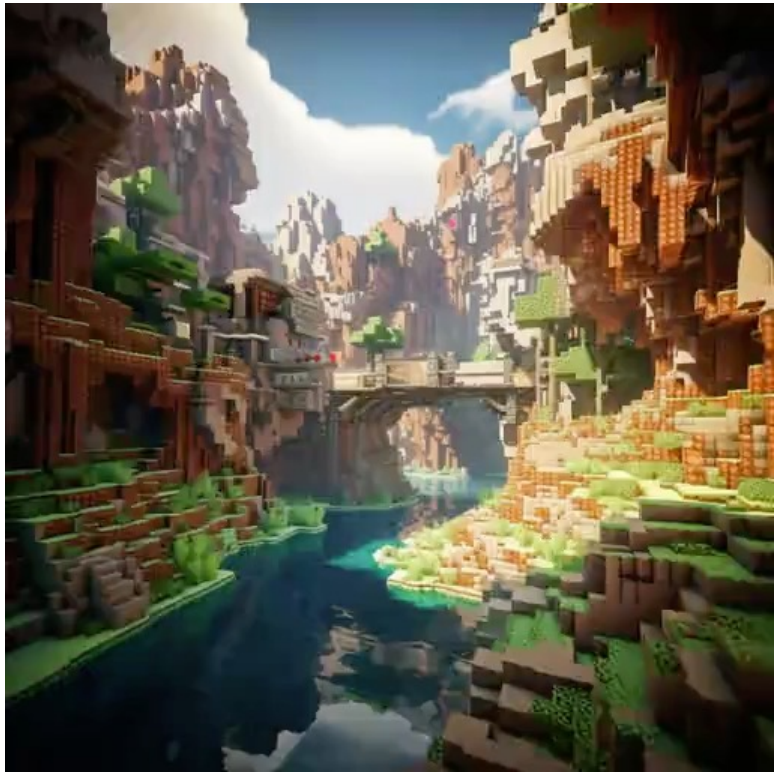
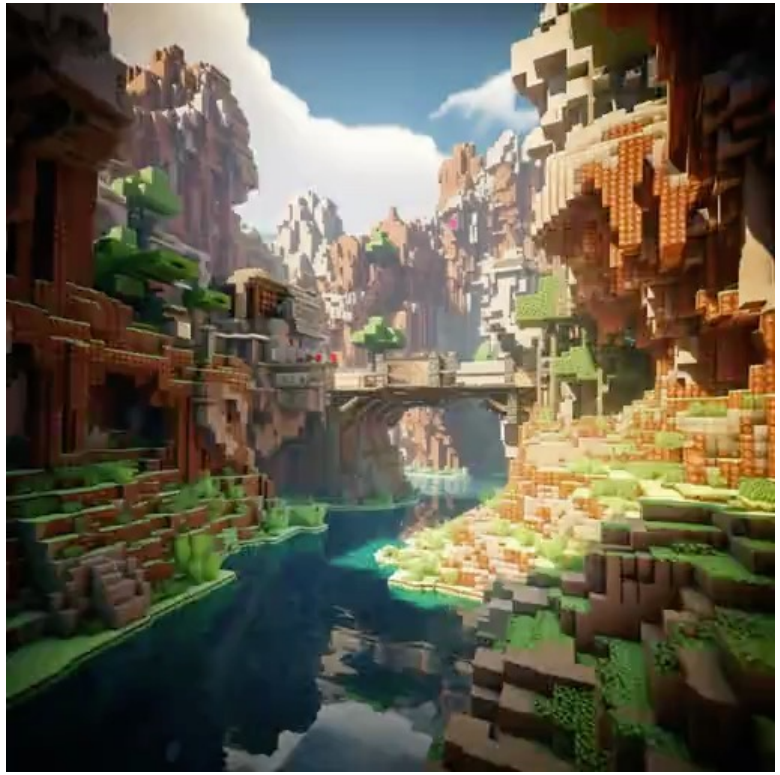






Results: To everywhere

Results: To everywhere



Results: Controlled wonderjourneys

千山鸟飞绝，
万径人踪灭。
孤舟蓑笠翁，
独钓寒江雪。

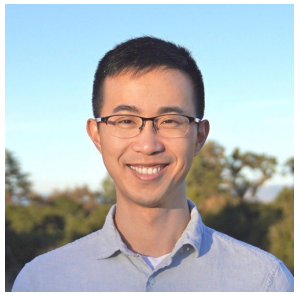
Results: Controlled wonderjourneys

Walden:

Thoreau's arrival...
Self-sufficiency...
Pond in winter...

WonderJourney:

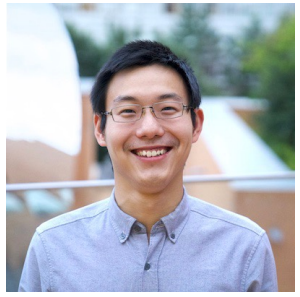
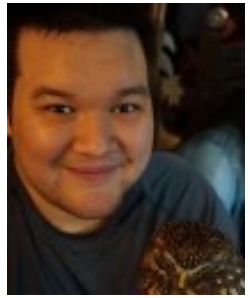
Going from Anywhere to Everywhere









kovenyu.com/wonderjourney/